

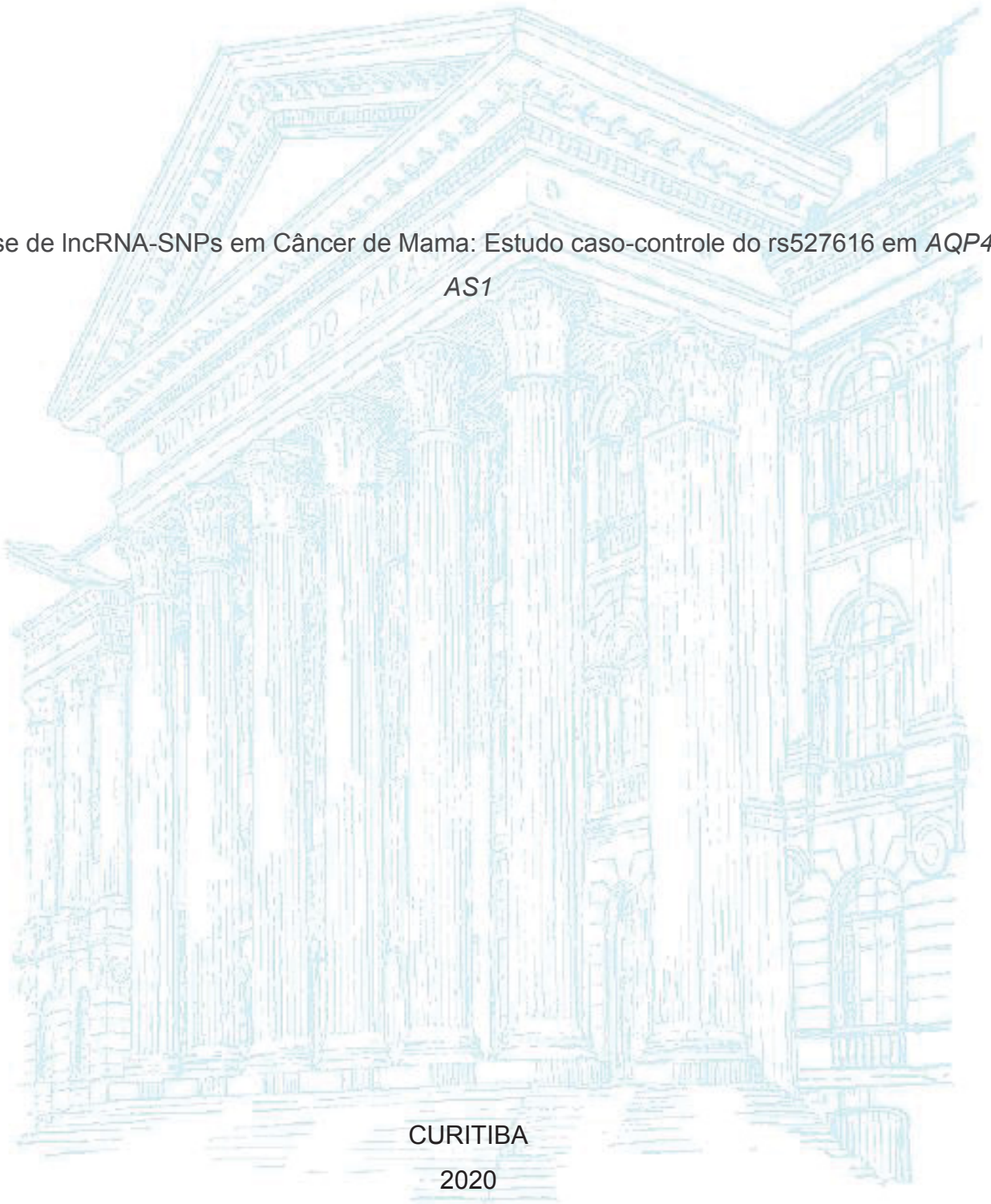
UNIVERSIDADE FEDERAL DO PARANÁ

RAFAEL DIOGO MARCHI

Análise de lncRNA-SNPs em Câncer de Mama: Estudo caso-controle do rs527616 em *AQP4-AS1*

CURITIBA

2020



RAFAEL DIOGO MARCHI

Análise de lncRNA-SNPs em Câncer de Mama: Estudo caso-controle do rs527616 em *AQP4-AS1*

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética do Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná.

Orientadora: Prof^a. Dr^a. Jaqueline Carvalho de Oliveira

Co-Orientador: Prof. Dr. Iglénir João Cavalli

CURITIBA

2020

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Rosilei Vilas Boas – CRB/9-939).

Marchi, Rafael Diogo.

Análise de lncRNA-SNPs em câncer de mama: estudo caso-controle do rs527616 em AQP4-AS1. / Rafael Diogo Marchi. – Curitiba, 2020.
64 f. : il.

Orientadora: Jaqueline Carvalho de Oliveira.

Coorientador: Iglenir João Cavalli.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Mamas - Câncer. 2. Mulheres – Curitiba (PR). 3. Mamas - Tumores. 4. Polimorfismo (Genética). I. Título. II. Oliveira, Jaqueline Carvalho de. III. Cavalli, Iglenir João. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 615.82



MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -
40001016006P1

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de ¹Mestrado de **RAFAEL DIOGO MARCHI** intitulada: **Análise de lncRNA-SNPs em câncer de mama: Estudo caso-controle do rs527616 em AQP4-AS1**, sob orientação da Profa. Dra. JAQUELINE CARVALHO DE OLIVEIRA, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 24 de Março de 2020.

Assinatura Eletrônica
03/04/2020 10:01:42.0
DANIELA FIORI GRADIA
Presidente da Banca Examinadora

Assinatura Eletrônica
03/04/2020 10:37:51.0
RICARDO LEHTONEN RODRIGUES DE SOUZA
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
20/05/2020 14:00:35.0
IGLENIR JOÃO CAVALLI
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
03/04/2020 15:32:22.0
JAQUELINE CARVALHO DE OLIVEIRA
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
03/04/2020 22:03:52.0
EDNEIA AMANCIO DE SOUZA RAMOS CAVALIERI
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

AGRADECIMENTOS

A Prof. Dra. Jaqueline Carvalho de Oliveira pela incrível orientação e paciência, por me tranquilizar em momentos de dificuldades e por tudo que me ensinou nesses 2 anos de orientação.

Ao Prof. Dr. Iglénir João Cavalli por ter aceitado fazer parte desse estudo como coorientador, estar sempre disposto a ajudar e por ser um incrível exemplo de pesquisador e pessoa.

Aos membros da banca examinadora, Professores Drs. Ricardo Lehtonen Rodrigues de Souza e Edneia Amancio De Souza Ramos Cavalieri e Marcia Holsbach Beltrame por aceitarem analisar este trabalho.

Aos membros da banca de acompanhamento, Prof. Dra. Lupe Furtado Alle e Prof. Dra. Marcia Holsbach Beltrame, por todas as contribuições que fizeram neste trabalho ao longo destes 2 anos.

Aos colegas de Laboratório de Citogenética Humana e Oncogenética por toda a ajuda, companheirismo e paciência.

A todas as pacientes envolvidas neste estudo.

À CAPES pelo apoio financeiro.

Em especial...

Aos meus pais, Valdir Marchi e Alaide Marchi, por serem exemplos de pai e mãe, por terem me apoiado e me tranquilizado nos momentos mais difíceis.

A minha avó Paula Baldo e Tia Alair Fronza, por terem me ajudado financeira nos momentos de dificuldade.

A meus queridos amigos Guillermo, Gabriel, Leandro, Tamyres, Tayane, Ana Flavia, Ana Carolina, Jéssica, Rafaela e Erika por estarem ao meu lado durante todo o decorrer desse trabalho.

A Ana Luiza Mattana, por estar sempre pronta a me ajudar, pelas conversar e por ser um exemplo de pessoa.

A Doutoranda Carolina Mathias por ter me ajudada no início desta caminhada e ter me ensinado muito sobre todos os experimentos realizados em laboratório.

À minha incrível esposa Paula Cristine Vilar Sarter, por ter mudado minha vida, ter feito desse sonho uma realidade, ter me aguentado por anos e me erguido em meus piores momentos.

RESUMO

O câncer de mama é a neoplasia mais comum em mulheres em todo o mundo e a principal causa de morte por câncer entre as mulheres. Sua incidência e mortalidade podem variar consideravelmente pela influência de muitos fatores complexos. Entre os fatores envolvidos na tumorigênese estão os RNAs longos não codificantes (lncRNAs), RNAs funcionalmente ativos e com mais de 200 nucleotídeos. A diferença de expressão de muitos lncRNAs tem sido associado com diversos tipos tumorais. Diferenças de expressão gênica muitas vezes podem ser desencadeadas por variações no DNA, entre essas, os polimorfismos de nucleotídeo único (Single Nucleotide Polymorphism – SNPs). Estudos de associação genômica em larga escala, como, por exemplo, Genome-wide Association Studies (GWAS) demonstraram que a maioria dessas variantes são encontradas em regiões não codificantes do genoma, das quais incluem os lncRNAs. O objetivo principal deste trabalho foi avaliar lncRNAs diferencialmente expressos e contendo SNPs associados com o câncer de mama em dados públicos de larga escala e avaliar especificamente o polimorfismo rs527616, localizado no lncRNA *AQP4-AS1*, em 306 pacientes portadoras de câncer de mama e em 312 controles sem histórico de neoplasias, provenientes de uma população brasileira da cidade de Curitiba, PR. A partir da análise de dados brutos de 12.000 lncRNAs (TCGA), obtivemos 1334 lncRNAs diferencialmente expressos em câncer de mama comparados a tecido normal. Desses, 22 lncRNAs continham em suas sequências SNPs previamente associados em estudo GWA (*Genome wide association*). Para o rs527616, as genotipagens foram realizadas utilizando a técnica de PCR-SSP. A comparação das frequências genotípicas entre pacientes e controles apresentou associação com a suscetibilidade com o câncer de mama quando comparados os genótipos GGxGG (OR= 1,38; p= 0,035; IC95% 1,01-1,90) e CGXCG (OR= 1,70; p= 0,0009; IC95% 1,23-2,34), ao comparar os genótipos CCXCC foi observado um valor de OR=0,34; p=0,004; IC95% 0,15-0,75, conferindo fator protetivo ao desenvolvimento do câncer de mama. A relação entre os alelos CXG, não apresentou valor estatístico significativo. O Qui-quadrado de homogeneidade não demonstrou diferenças entre os subtipos luminais/triplo negativo, invasão de linfonodos regionais e grau de diferenciação tumoral. Neste trabalho mostramos de forma inédita um estudo caso-controle do rs527616 em uma amostra de população brasileira. Com base nestes resultados, concluímos que, nesta amostra, os genótipos GG e CG sugerem uma possível associação a suscetibilidade ao câncer de mama, assim como o genótipo CC apresenta uma associação protetiva ao desenvolvimento de câncer de mama.

Palavras-chave: rs527616; lncRNA *AQP4-AS1*; Câncer de mama; População brasileira.

ABSTRACT

Breast cancer is the most common cancer in women worldwide and the leading cause of cancer death among women. Its incidence and mortality can vary considerably due to the influence of many complex factors. Among the factors involved in tumorigenesis are long non-coding RNAs (lncRNAs), functionally active RNAs with more than 200 nucleotides. The difference in expression of many lncRNAs has been associated with several tumor types. Differences in gene expression can often be triggered by variations in DNA, including single nucleotide polymorphisms (Single Nucleotide Polymorphism - SNPs). Large-scale genomic association studies, such as, for example, Genome-wide Association Studies (GWAS) have shown that most of these variants are found in non-coding regions of the genome, of which include lncRNAs. The main objective of this work was to evaluate differentially expressed lncRNAs and containing SNPs associated with breast cancer in large-scale public data and to specifically evaluate the rs527616 polymorphism, located in the lncRNA *AQP4-AS1*, in 306 patients with breast cancer and in 312 controls with no history of neoplasms, from a Brazilian population in the city of Curitiba, PR. From the analysis of raw data of 12,000 lncRNAs (TCGA), we obtained 1334 lncRNAs differentially expressed in breast cancer compared to normal tissue. Of these, 22 lncRNAs contained SNPs in their sequences previously associated in a GWA (Genome wide association) study. For rs527616, genotyping was performed using the PCR-SSP technique. The comparison of genotypic frequencies between patients and controls was associated with susceptibility to breast cancer when the GGxGG genotypes (OR = 1.38; $p = 0.035$; 95% CI 1.01-1.90) and CGXCG (OR = 1.70; $p = 0.0009$; 95% CI 1.23-2.34), when comparing the CCXCC genotypes, was observed an OR value of 0.34; $p = 0.004$; 95% CI 0.15-0.75, conferring a protective factor to the development of breast cancer. The relationship between the CXG alleles, did not present significant statistical value. The homogeneity chi-square did not show differences between luminals/triple negative subtypes, invasion of regional lymph nodes and degree of tumor differentiation. In this work we show an unprecedented case-control study of rs527616 in a sample of the Brazilian population. Based on these results, we conclude that, in this sample, the GG and CG genotypes suggest a possible association with susceptibility to breast cancer, just as the CC genotype has a protective association with the development of breast cancer.

Keywords: rs527616; lncRNA *AQP4-AS1*; Breast cancer; Brazilian population.

LISTA DE FIGURAS

FIGURA 1. Estrutura anatômica da mama, carcinoma <i>in situ</i> e invasivo	16
FIGURA 2. Classificação dos lncRNAs	21
FIGURA 3. Ações realizadas por lncRNAs	22
FIGURA 4. Aquaporinas e suas funções em diferentes tipos de câncer	25
FIGURA 5. Sequência Fasta da fita antisense e <i>primers</i> do lncRNA <i>AQP4-AS1</i> , em verde <i>primers</i> para amplificação do alelo G, em amarelo <i>primers</i> para amplificação do alelo C	33
FIGURA 6. Padrão eletroforético dos fragmentos amplificados	34
FIGURA 7. HeatMap de lncRNA diferencialmente expressos em câncer de mama.....	37
FIGURA 8. Diferença de expressão dos lncRNA nos subtipos moleculares de câncer de mama. Destacado em vermelho o lncRNA analisado neste trabalho	39
FIGURA 9. Diferença de expressão dos lncRNA nos subtipos moleculares de câncer de mama	40
FIGURA 10. Expressão do <i>AQP4-AS1</i> em diferentes subtipos de câncer de mama	42
FIGURA 11. Sequenciamento do lncRNA <i>AQP4-AS1</i> , amostra de genótipo GG	42
FIGURA 12. Sequenciamento do lncRNA <i>AQP4-AS1</i> , amostra de genótipo CC	43

LISTA DE FIGURAS

TABELA 1. Classificação do câncer de mama de acordo com características imunoistoquímicas	19
TABELA 2. Dados Clínicos e Histopatológicos de pacientes com Câncer de mama	30
TABELA 3. Primers utilizados para realização de PCRs alelo específicas na genotipagem das amostras de pacientes com câncer de mama e amostras controles	33
TABELA 4- Ciclagem da reação do lncRNA <i>AQP4-AS1</i>	34
TABELA 5- Primers utilizados no Sequenciamento do método Sanger	35
TABELA 6- Ciclagem da reação de sequenciamento realizada no aparelho MasterCycler Gradient, Eppendorf	36
TABELA 7- LncRNAs selecionados e seus respectivos dados	41
TABELA 8. Frequência genotípica e alélica dos pacientes e controles	44
TABELA 9- Frequência do SNP rs527616 em diferentes estudos	44
TABELA 10. Análises estatísticas de Odds Ratio	45
TABELA 11. Análise de Qui-Quadrado de homogeneidade para subtipos Imunoistoquímicos	46
TABELA 12. Análise de Qui-Quadrado de homogeneidade para metástase nos linfonodos e Grau de diferenciação tumoral	46

LISTA DE LISTA DE ABREVIATURAS E SIGLAS

AQPs – Aquaporinas

AQP0-AQP12 – diferentes Aquaporinas 0 até 12

AQP4-AS1 – Aquaporina 4 antisenso 1

CASC16 – Cancer susceptibility 16

CK5 – Citoqueratinas de padrão basal

dNTPs – Desoxinucleotídeo tri-fosfato

ddNTPs – Dideoxinucleotídeos tri-fosfato

EGFR – Receptor de crescimento epidermal, do inglês Epidermal Growth Factor Receptor

FGF-2 – Fator de crescimento de fibroblastos-2

HER2 – Human Epidermal Growth Factor Receptor 2 (Receptor de Fator de Crescimento epidermal 2)

HNSG – Hospital Nossa Senhora das Graças

HOTAIR – HOX transcript antisense intergenic RNA

INCA – Instituto Nacional de Câncer

LabCHO – Laboratório de Citogenética Humana e Oncogenética

LIGH – Laboratório de Imunogenética e Histocompatibilidade

LncRNAs – RNAs Longos Não Codificantes

LINC-PINT– Long intergenic non-protein coding RNA, p53 induced transcript

LINC00511 – Long intergenic non-protein coding RNA 511

LINC00536 – Long intergenic non-protein coding RNA 536

LINC02224 – Long intergenic non-protein coding RNA 2224

LINC01977 – Long intergenic non-protein coding RNA 1977

MAF – Minor allele frequency (Alelo de menor frequência populacional)

MEG3 – Maternally expressed 3

MIR4435-2HG – MIR4435-2 host gene

miRNAs – MicroRNAs

NCBI – National Center for Biotechnology Information

NCI – National Cancer Institute

ncRNAs – RNAs não codificantes

NHGRI – National Human Genome Research Institute

NPA – Região Asn-Pro-Ala

PAM50 – Prediction Analysis of Microarray 50

Pb – Pares de bases

PCR – Polymerase chain reaction (reação em cadeia da polimerase)

PR – Paraná

RE – Receptor de Estrogênio

RNAseq – sequenciamento de RNA

RP – Receptor de Progesterona

SNC – Sistema nervoso central

SNP – Single Nucleotide Polymorphism

TANRIC – The Atlas of non-coding RNA in Cancer

TCGA – The Cancer Genome Atlas

TOX3 – TOX high mobility group box family member 3

UFPR – Universidade Federal do Paraná

LISTA DE SÍMBOLOS

ml	Mililitro.
mg	Miligramma.
μg	Micrograma
μl	Microlitros
rpm	Rotações por minuto
χ^2	Qui-quadrado

SUMÁRIO

1. INTRODUÇÃO	15
2. REVISÃO DA LITERATURA	17
2.1 Câncer de Mama	17
2.2 RNAs Longos Não Codificantes	21
2.3 AQP4-AS1	26
3. OBJETIVOS	30
3.1 Objetivo Geral	30
3.2 Objetivos Específicos	30
4. MATERIAL E MÉTODOS	31
4.1 Análise de Bioinformática	31
4.2 Casuística	31
4.3 Análise de polimorfismos	33
4.3.1 Extração de DNA	33
4.3.2 PCR alelo específica	34
4.3.3 Sequenciamento pelo Método Sanger	37
4.4 Análise estatística	35
5. RESULTADOS	39
5.1 LncRNAs Diferencialmente Expressos em Câncer	39
5.2 LncRNAs com SNPs associados com Câncer de Mama	40
5.3 Genotipagem dos SNPs	45
5.4 Frequências Alélicas e Genotípicas	46
5.5 Estudo de Associação	47
5.6 Análise dos Subtipos e Parâmetros Histopatológicos	48
6. DISCUSSÃO	49
7. CONCLUSÃO	55
8. REFERÊNCIAS	56

1. INTRODUÇÃO

O câncer de mama é a quinta causa de morte no mundo, além de ser a neoplasia mais incidente em mulheres, excluindo tumores de pele não melanoma (GLOBOCAN, 2018). Dados divulgados pelo Instituto Nacional de Câncer (INCA) apontaram que no ano de 2018 foram computadas mais de 14 mil mortes causadas por essa neoplasia. Já para o ano de 2020 a estimativa é de que ocorram 66.280 novos casos de câncer de mama no Brasil (INCA, 2020).

Por se tratar de uma neoplasia heterogênea, sua classificação é dada por meio de características histopatológicas e moleculares, que desempenham um importante papel na determinação do diagnóstico, prognóstico e futuros tratamentos. A subdivisão molecular do câncer de mama, estabelecida inicialmente por Perou *et al.* (2000) e Sorlie *et al.* (2003), atualmente utiliza um conjunto de genes mínimos (PAM50), estabelecendo os seguintes subtipos: luminal A, luminal B, basal-*like*, HER2-positivo e normal-*like* (HENG *et al.*, 2017).

A prática de dividir os casos de câncer em grupos proporcionou um avanço significativo na melhora da taxa de sobrevida dos pacientes, porém, sua implementação nos sistemas de saúde é limitada, principalmente pelo seu alto custo (EROLES *et al.*, 2012). Dessa forma, uma classificação partindo de características clínico-patológicas e utilizando análises da expressão proteica de marcadores (imunoistoquímicas), como receptor de estrógeno (RE) e progesterona (RP), além de HER2 e Ki-67, foi estabelecida. Apesar do grande número de sistemas de classificação e da identificação de subtipos moleculares clinicamente relevantes, ainda existe uma alta taxa de mortalidade, o que justifica a busca por novos marcadores de prognósticos, entre os quais os RNAs não codificantes se destacam.

Sabe-se, que cerca de 98% de toda a produção transcricional em humanos é RNA não codificante (ncRNAs) e que existem evidências de suas participações em diversos processos celulares, sugerindo que eles exercem um papel essencial na programação genômica em humanos e em outros organismos complexos

(MATTICK, 2001; 2009). Entre os ncRNAs, destacam-se os longos não codificantes (lncRNAs) que apresentam mais de 200 nucleotídeos.

lncRNAs abrangem um amplo espectro de funções, principalmente relacionados com a regulação da expressão gênica. Do ponto de vista da sua regulação, compartilham semelhanças com transcritos codificantes, podendo ser transcritos pela RNA polimerase II, serem poliadenilados, sofrerem *splicing* e também serem submetidos a modificadores epigenéticos, tais como regulação por metilação de DNA e modificações de histonas (MATTICK, 2009). Apesar da maioria dos lncRNAs ainda não terem função conhecida, o número de longos com funções descritas está aumentando constantemente.

Neste trabalho, inicialmente analisamos os dados de expressão de 3600 lncRNAs em 837 amostras tumorais e 105 tecidos mamários não tumorais de pacientes com câncer de mama através de dados de *RNAseq* depositados no TCGA. Posteriormente, avaliamos no GWAS CATALOG, quais desses lncRNAs possuíam SNPs associados com câncer de mama. Dos SNPs evidenciados e mapeados em lncRNAs diferencialmente expressos no câncer de mama, selecionamos o rs527616 (C>G) localizado no *AQP4-AS1*, para análise em um segundo grupo de pacientes. Esse lncRNA está na fita antisenso do gene *AQP4* já descrito previamente como sendo importante no câncer de mama, porém, não há nenhum estudo evidenciando a importância especificamente desse lncRNA.

O SNP rs527616 já foi estudado em estudos de larga escala em algumas populações, porém, nenhum estudo foi realizado com população de origem brasileira. Para investigar o efeito deste SNP no câncer de mama em uma amostra da população do sul do Brasil, realizamos um estudo de associação do tipo caso-controle com 306 pacientes de câncer de mama e 312 controles, ambas originárias de Curitiba, PR.

2. REVISÃO DA LITERATURA

2.1 Câncer de Mama

O câncer de mama é a neoplasia mais comum em mulheres, excluindo tumores de pele não melanoma (FERLAY, 2015), e a quinta causa de mortes (6,6%) comparada a todos os outros tipos de cânceres (GLOBOCAN, 2018).

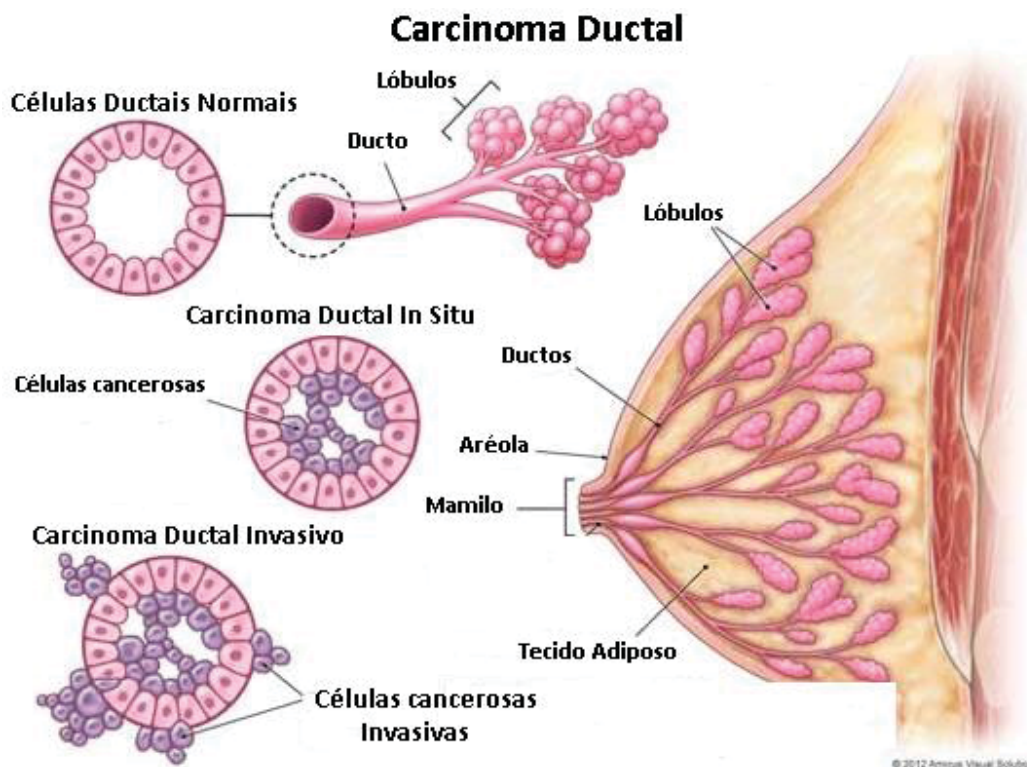
Dados divulgados pelo Instituto Nacional de Câncer (INCA) mostraram que em 2018 houveram mais de 14 mil mortes por câncer de mama no Brasil e é estimado cerca de 66.280 novos casos de câncer de mama para o ano de 2020 foi estimado, podendo ultrapassar 16 mil mortes (INCA, 2020).

O câncer de mama consiste em um grupo heterogêneo de doenças, que apresentam características histopatológicas e moleculares distintas, que são utilizadas na determinação do diagnóstico, prognóstico e futuros tratamentos. Sendo assim, diferentes tipos de classificação são utilizados.

Aproximadamente 95% dos casos de tumor mamário são carcinomas, que são tumores derivados de células epiteliais (MAKKI, 2015). Inicialmente, o carcinoma de mama é classificado pelo seu local de origem: carcinomas ductais, quando presentes nos ductos, ou lobulares, quando encontrados nos lóbulos da mama (MAKKI, 2015) (FIGURA 1).

Adicionalmente, também é analisada a capacidade migratória dessa neoplasia. São classificados como carcinomas *in situ*, quando o crescimento celular está restrito apenas a seu tecido de origem, o que representa um melhor prognóstico. Já os carcinomas invasivos, têm maior capacidade de invadir outros tecidos e se espalharem para outros órgãos. Nesse caso, apresentam um pior prognóstico (BATEMAN; SHAW, 2016).

FIGURA 1. Estrutura anatômica da mama, carcinoma in situ e invasivo.



FONTE: Adaptado de Amicus Visual Solutions. Disponível: https://www.anatomyexhibits.com/downloads/12199_02x/.

A prática de dividir os casos de câncer em grupos possibilitou o melhor encaminhamento de cada paciente para o devido tratamento, de acordo com o prognóstico. Essa prática levou a uma significativa melhora na taxa de sobrevivência dos pacientes (EROLES *et al.*, 2012).

De acordo com o sistema Nottingham Histologic Score (modificado de Elston-Ellis do sistema de graduação de Scarff-Bloom-Richardson) podemos dividir os cânceres de mama utilizando três fatores de diferenciação celular: quantidade de formação de glândulas, pleomorfismo e atividade mitótica. Assim, analisando essas características histológicas é possível graduar os cânceres de mama de acordo com sua agressividade, como segue: tumores classificados com grau I (baixo) caracterizam-se por serem bem diferenciados e apresentam melhor prognóstico;

grau II (intermediário) moderadamente diferenciados e grau III (alto), pouco diferenciados e de pior prognóstico.

Outro sistema utilizado para a classificação dos tumores malignos é o sistema TNM, em que T significa a extensão do tumor primário, N é a ausência ou presença e extensão da metástase nos linfonodos regionais e M representa a ausência ou presença de metástase à distância (WILEY-BLACKWELL, 2009).

Levando em conta a heterogeneidade do câncer de mama, além dos modelos de classificação anteriormente citados, são também utilizados na prática clínica biomarcadores, como receptores de membranas e padrões de expressão gênica. A melhor classificação disponível até o momento, é a classificação molecular do câncer de mama. Essa classificação teve início com o trabalho pioneiro de Perou e colaboradores, em 2000, e foi validado por Sorlie e colaboradores em 2003, no qual foi analisando o perfil de expressão gênica por tecnologia de microarranjo e descritos 5 subtipos diferentes dessa neoplasia, sendo eles: luminal A, luminal B, HER2-*overexpressing*, basal-*like* e normal-*like*. Posteriormente, outros grupos foram incluídos, como: claudina baixa e molecular apócrino (PRAT, *et al.*, 2010).

Atualmente, com base na análise de assinatura molecular e focando em um conjunto de genes mínimos (PAM50, do inglês *Prediction Analysis of Microarray* 50), com maior aplicação clínica, pacientes diagnosticados com câncer de mama são classificados em cinco subtipos principais: luminal A, luminal B, basal-*like*, HER2-*enriched* e Normal-*like* (HENG, 2017). Cada subtipo apresenta características biológicas e epidemiológicas específicas que auxiliam no prognóstico clínico de cada paciente.

O subtipo luminal A é o mais comum entre os cinco, representando 50-60% do total de ocorrências. Seu perfil imunoistoquímico é caracterizado por expressar positivamente ER e PR, HER2 negativo e baixa taxa de proliferação celular (EROLES *et al.*, 2012).

Comparado ao luminal A, o subtipo luminal B possui fenótipo mais agressivo. Constitui de 10-20% das ocorrências e apresenta perfil imunoistoquímico com ER

e/ou PR positivo, HER2 positivo ou negativo com maior índice de proliferação celular (EROLES *et al.*, 2012).

Tumores classificados como *HER2-enriched* correspondem de 15-20% dos casos de cânceres de mama, com alta expressão de HER2 e de genes associados à sua via, além de possuírem ER e PR negativos. Morfologicamente esse subtipo possui alto índice de proliferação celular com pior prognóstico, e em 75% dos casos apresentam alto grau histológico (EROLES *et al.*, 2012).

O subtipo *basal-like* representa de 10-20% de todas as ocorrências de carcinomas mamários possuindo ER, PR e HER2 negativos e *CK5* (Citoqueratinas de padrão basal) e *EGFR* (Receptor de fator de crescimento epidérmico) positivos. Comparado com os luminais, esse subtipo apresenta pior prognóstico, alto índice mitótico e necrose (EROLES *et al.*, 2012).

Normal-like representa de 5-10% das ocorrências. Esse subtipo possui o perfil imunoistoquímico mais controverso dos 5 classificados por apresentar ER, PR e HER2 negativos. Contudo, não pode ser classificado como *Basal-like* por também possuir *CK5* e *EGFR* negativos. Esse subtipo apresenta prognóstico intermediário ao luminal e *basal-like* e expressa genes característicos de tecidos adiposos (EROLES *et al.*, 2012).

Embora a classificação molecular seja de extrema importância para o prognóstico das pacientes, a implementação desta nos sistemas de saúde não é algo viável a curto e médio prazo. Dessa forma, foi sugerida uma classificação que leva em consideração características clinico-patológicas, que são estratificadas de acordo com a análise imunoistoquímica dos tumores, uma vez que este método é mais viável para a implementação nos sistemas de saúde pública. Por meio dessa análise é possível realizar uma aproximação dos subtipos moleculares, de acordo com a expressão proteica dos seguintes marcadores: receptor de estrógeno (ER) e progesterona (PR), HER2 e o marcador de proliferação celular Ki-67 (TABELA 1).

TABELA 1. Classificação do câncer de mama de acordo com características imunoistoquímicas.

SUBTIPO	RE/RP/HER2	Ki-67	PROGNÓSTICO
Luminal A	RE+RP+HER2-	Baixo <14%	Bom
Luminal B	RE+RP+/-HER2+	Baixo <14%	Intermediário
	RE+RP-HER2-	Alto ≥14%	Desfavorável
HER2-positivo	RE-RP-HER2+	Alto ≥14%	Desfavorável
Triplo-negativo	RE-RP-HER2-	Alto ≥14%	Desfavorável

LEGENDA: RE, receptor de estrogênio, RP, receptor de progesterona, HER2, proteína codificada pelo oncogene ERBB2/HER, Ki-67 marcador de proliferação celular. FONTE: Adaptado de GOLDBIRSCH *et al.* (2013).

Apesar do grande número de sistemas de classificação e da identificação de subtipos moleculares clinicamente relevantes, a classificação e o tratamento atual dos pacientes permanecem deficitários em muitos centros e ainda existe uma alta taxa de mortalidade para a doença, o que justifica a busca de novos marcadores de prognóstico. Nesse contexto, o estudo de RNAs não codificantes é promissor.

2.2 RNAs Longos Não Codificantes

Cerca de 98% de toda a produção transcricional em humanos são RNAs não codificantes (ncRNAs). Há evidências crescentes de que esses transcritos são funcionais e participam em inúmeros processos celulares, particularmente na regulação da expressão gênica e de processos epigenéticos em humanos e em outros organismos complexos (CALLE, *et al.*, 2018).

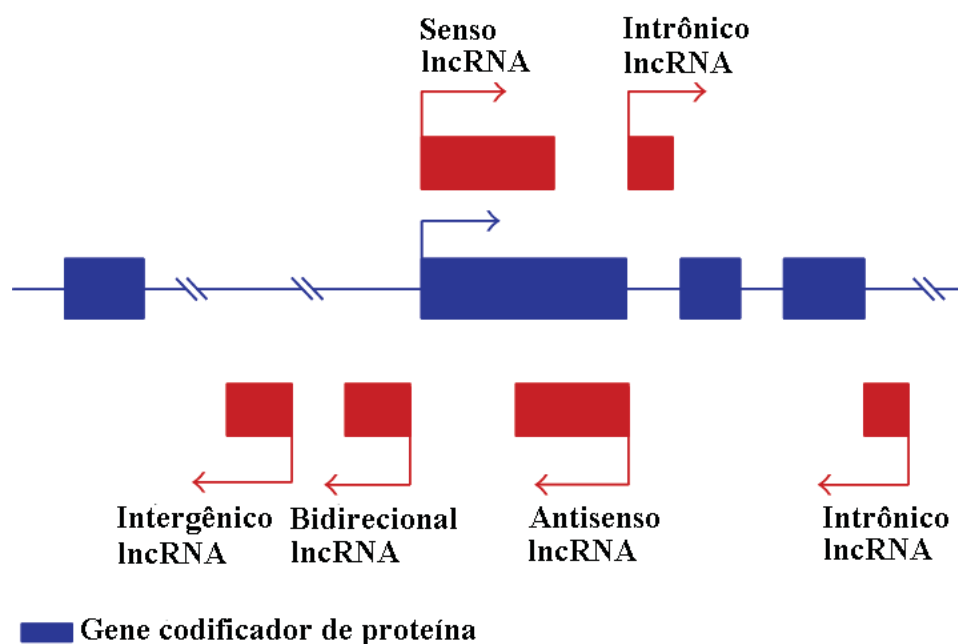
Os ncRNAs são geralmente divididos em duas categorias principais, de acordo com seu tamanho: ncRNAs pequenos, transcritos que apresentam menos

de 200 nucleotídeos de comprimento e ncRNAs longos (lncRNAs), maiores que 200 nucleotídeos. Pequenos ncRNAs incluem os bem caracterizados microRNAs (miRNAs), que atuam como importantes reguladores pós-transcricionais da expressão gênica via complexo silenciador induzido por RNA. Por outro lado, a classe de lncRNAs é menos compreendida e parece estar envolvida em diversos processos biológicos, como na remodelagem da cromatina, modificações de histonas e regulação da expressão gênica, podendo atuar em nível transcricional e pós-transcricional (KWOK; TAY, 2017).

Os lncRNAs compartilham algumas semelhanças com transcritos codificantes: eles podem ser transcritos pela RNA polimerase II, apresentar íntrons e variantes de *splicing*, podem exibir cauda poli-A na extremidade 3' e podem ter sua transcrição regulada por modificadores epigenéticos semelhantes aos genes codificadores de proteínas, tais como regulação por metilação de DNA e modificações de histonas (MATTICK, 2009).

LncRNAs abrangem várias moléculas de RNA diferentes e, por seu amplo espectro de funções, é difícil classificá-los, porém, sua classificação tradicional é baseada na localização com relação a de genes codificantes próximos, permitindo a descrição em cinco categorias: antisense, intrônico, bidirecional, intergênico e senso (WU; DU, 2017) (FIGURA 2).

FIGURA 2. Classificação dos lncRNAs.

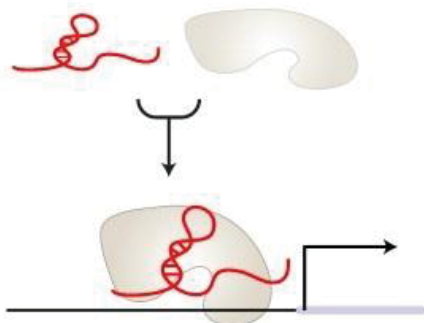
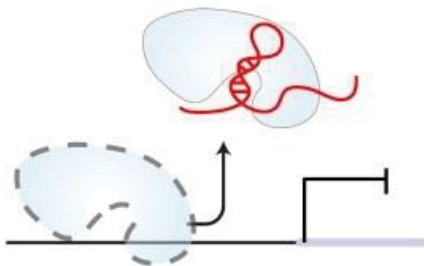
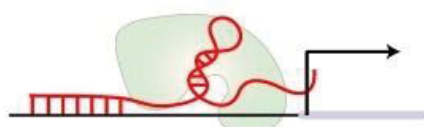
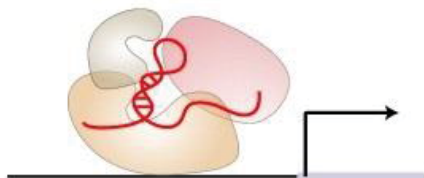
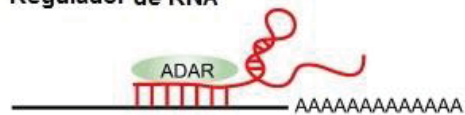
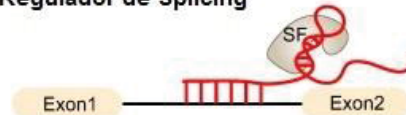
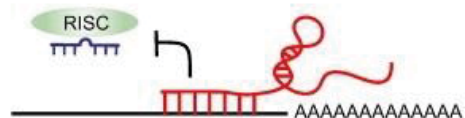
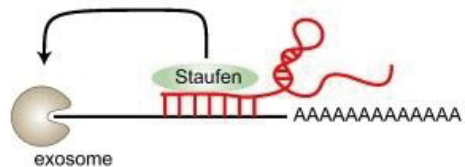
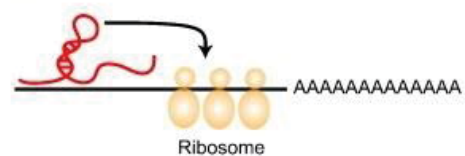


FONTE: adaptado de LOSKO, M; KOTLINOWSKI, J; JURA, J, 2016.

Apesar da maioria dos lncRNAs ainda não ter função conhecida, o número de lncRNAs com funções descritas está aumentando constantemente, muitas dessas funções associadas com sua capacidade de regulação da expressão gênica (GEISLER; COLLIER, 2013).

Das ações conhecidas, os lncRNAs possuem a capacidade de interagir diretamente com fatores e repressores transcricionais, além de agirem como iscas carregando fatores de transcrição para longe da cromatina. Podem, também, participar do *splicing* alternativo, abrigar precursores de miRNA e atuar como guias transcricionais, agindo como arcabouços (*scaffold*) para múltiplas proteínas que formam complexos de ribonucleoproteínas. Adicionalmente, lncRNAs podem abrigar o local de reconhecimento de miRNAs funcionais, direcionar a degradação do RNA e modular positiva ou negativamente a eficiência da tradução (YANG; LU; YUAN, 2014) (FIGURA 3).

FIGURA 3. Ações realizadas por lncRNAs.

Epigenético e Transcricional**a Ativador Transcricional****b Repressor Transcricional****c Guia Transcricional****d Arcabouço para Complexos Modificadores de Cromatina****Pós-Transcricional****e Regulador de RNA****f Regulador de Splicing****g Ancorador de miRNA****h Sequestrador de miRNA****i Bloqueador de miRNA****j Regulador na degradação do RNA****k Regulador na Eficiência de Tradução**

FONTE: Adaptado de YANG; LU; YUAN, 2014.

Alterações no perfil de expressão dessas moléculas têm sido associadas com uma variedade de doenças, incluindo câncer (CIPOLLA *et al.*, 2018; OLIVEIRA *et al.*, 2019).

A identificação de lncRNAs importantes em vários processos da carcinogênese acrescenta informações para uma melhor compreensão da heterogeneidade tumoral. Considerando que lncRNAs parecem exibir um padrão de expressão tecido-específico, a avaliação da expressão de lncRNAs e de variações genéticas associadas com diferenças de expressão desses lncRNAs, na prática clínica, pode auxiliar na elaboração de melhores diagnósticos e subclassificações moleculares mais acuradas (WU; DU, 2017).

O *The Cancer Genome Atlas* (TCGA) é uma colaboração entre o *National Cancer Institute* (NCI) e o *National Human Genome Research Institute* (NHGRI) que gerou uma grande quantidade de dados. Este, contém informações em larga escala de diversas variações genômicas, transcriptômicas e epigenômicas em 33 tipos de câncer, dentre os quais está o câncer de mama. Nessa plataforma, os dados de expressão dos lncRNAs estão disponíveis para uso público (TCGA, 2018).

Diferenças de expressão gênica muitas vezes podem ser desencadeadas por variações no DNA, entre essas, os polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphism* – SNPs). Os SNPs são o tipo de polimorfismo mais frequente no genoma humano, com frequência genômica média de aproximadamente um por 1.000 pares de base (pb) (KIM; MISRA, 2007).

Estudos de associação genômica em larga escala, como, por exemplo, *Genome-wide Association Studies* (GWAS) demonstraram que a maioria dessas variantes são encontradas em regiões não codificantes do genoma.

GWAS é um delineamento experimental utilizado para detectar associações entre variantes genéticas e características de interesse em amostras populacionais (VISSCHER *et al.*, 2017). A realização desse delineamento foi possível por meio da união de muitos consórcios em todo o mundo, como o Consórcio do Câncer de Mama da Ásia (ABCC) e o Consórcio da Associação de Câncer de Mama (BCAC). Por conseguinte, os dados disponibilizados pelo GWAS CATALOG (2018), apontam

que, em 118 estudos realizados, foram catalogadas mais de 1500 associações de SNPs com câncer de mama.

2.3 AQP4-AS1

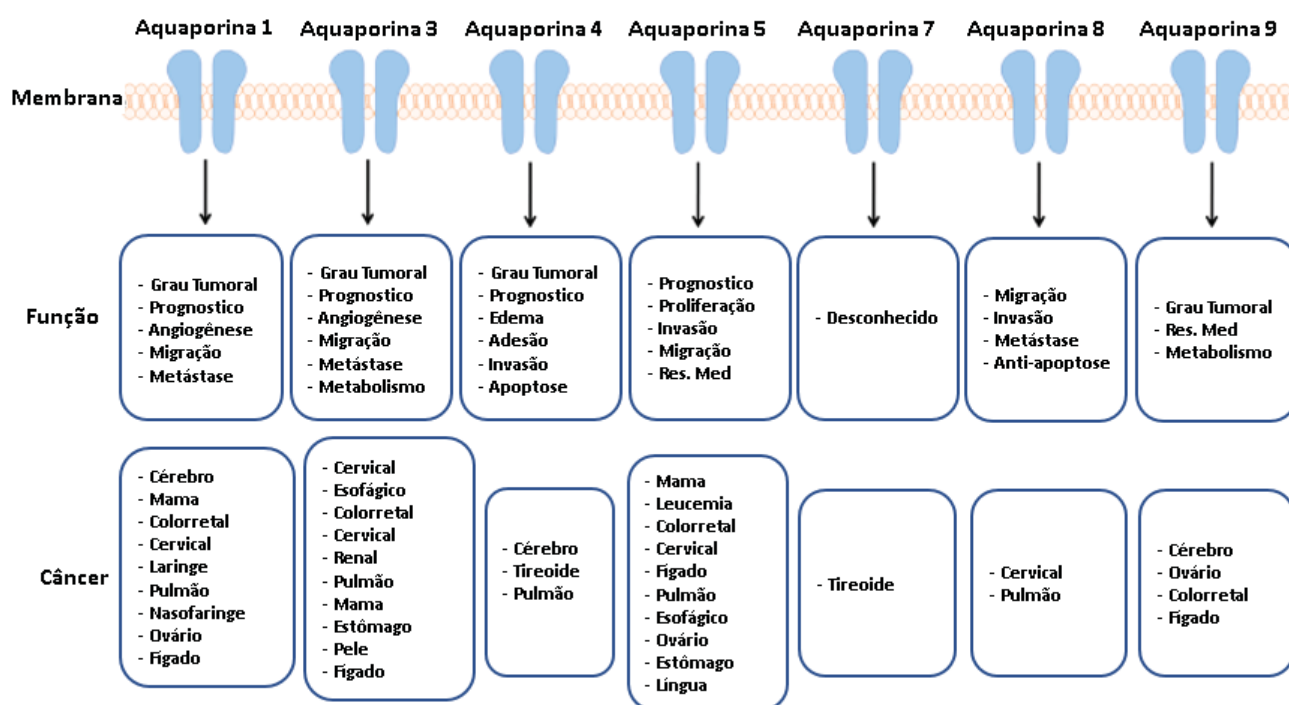
As aquaporinas (AQPs) são uma família de pequenas proteínas de transporte de membrana que se reúnem como tetrâmeros e que acredita-se serem expressas na maioria, se não, todas as membranas celulares e intracelulares, atuando principalmente como poros seletivos à água e sendo responsáveis por regular o movimento rápido da água através de barreiras epiteliais impulsionadas por gradientes osmóticos. (VERKMAN *et al*, 2008; MOBASHERI., BARRETT-JOLLEY, 2013).

Duas regiões do canal são críticas para a função AQP. Essas regiões são Asn-Pro-Ala (NPA) e a região aromática-arginina. A região da NPA fica localizada no meio do canal e participa na exclusão de prótons e ajuda a localizar AQPs na membrana plasmática. A região aromática-arginina está voltada para o lado extracelular e atua como um filtro de seletividade, bloqueando partículas maiores que a água (KREIDA, S; HORSEFIELD, S-H. 2015).

Até o momento, 13 membros da família de genes das aquaporinas foram identificados em humanos: AQP0-AQP12 (CASTLE. 2005). As proteínas codificadas pelos genes da aquaporina foram classificadas em dois grupos com base nas permeabilidades do substrato: 1) Aquaporinas clássicas permeáveis à água, que incluem as AQP1, AQP2, AQP4, AQP5 AQP6, AQP8 e AQP11; 2) As aquagliceroporinas permeáveis a água e uma variedade de pequenos solutos neutros, incluindo substâncias como glicerol e uréia, incluem as AQP3, AQP7, AQP9 e AQP10. (AGRE P, et al.; 2002; KREIDA, S; HORSEFIELD, S-H. 2015).

As aquaporinas desempenha um papel significativo no câncer, participando de processos como migração, invasão, metástase, proliferação, resistência a medicamentos e podem ser associadas ao prognóstico (FIGURA 4). A existência de associação entre vários tipos de câncer e canais AQP, sugerem que eles não desempenham um papel passivo apenas no transporte de água (DAJANI, S; *et al.*, 2018).

FIGURA 4. Aquaporinas e suas funções em diferentes tipos de câncer.



FONTE: adaptado de DAJANI, S; *et al.*, 2018.

Devido ao papel das AQPs, a alteração em sua expressão ou de suas funções no câncer, acredita-se que essas proteínas possam ter potencial terapêutico, assim como potencial de serem usadas como biomarcadores de prognóstico (TRADTRANTIP, L; *et al.*, 2017; DAJANI, S; *et al.*, 2018)

Algumas aplicações potencialmente interessantes já foram descritas, como os inibidores da AQP1, que sugerem ter atuação como diuréticos únicos, participando na inibição a angiogênese e crescimento tumoral e redução da pressão intraocular no glaucoma. Inibidores do transporte de água na AQP4 sugerem reduzir o inchaço cerebral no derrame isquêmico, já os inibidores do transporte de glicerol e ou peróxido de hidrogênio na AQP3 sugerem participação na prevenção ou retardo do crescimento de tumores cutâneos e doenças inflamatórias da pele (TRADTRANTIP, L; *et al.*, 2017).

A fim de atestar a utilidades das AQPs como alvos terapêuticos, TRADTRANTIP e colaboradores (2017), produziram um anticorpo anti-AQP4 de alta

afinidade (*aquaporumab*), com o qual observaram em culturas celulares o impedimento a citotoxicidade em células de pacientes com neuromielite óptica.

A aquaporina 4 (AQP4), um membro da família AQP, tem um papel fundamental na manutenção da homeostase da água, que se acredita estar associada ao desenvolvimento de tumores (LI; SUN; HAN, 2016).

A AQP4 tornou-se um alvo terapêutico interessante em vários distúrbios neurológicos, devido expressão generalizada no sistema nervoso, sendo a proteína do canal de água predominante no cérebro dos mamíferos e expressa em astrócitos por todo o sistema nervoso central, o que lhe permite controlar a homeostase da água do sistema nervoso através do seu envolvimento no movimento da água através da barreira hematoencefálica (VANDERBROEK, A; YASUI, M. 2020).

Embora tenha sido descrita apenas como canal específico da água, a perda de AQP4 foi associada a níveis alterados de K⁺ extracelular, sugerindo uma possível participação na homeostase do K⁺ (VERKMAN, A.S; BINDER. D.K; *et al.* 2006).

A AQP4 foi associada ao papel pró-inflamatório intrínseco no inchaço dos astrócitos e na liberação de citocinas, e uma redução no transporte de água pelo AQP4 parece ser protetora nas doenças neuroinflamatórias do SNC, como encefalomielite autoimune (EAE) (LI. L; *et al.*, 2011)

Estudos indicam que existe participação da AQP4 na invasão e migração de células em glioma humano, assim como, o silenciamento da AQP4 também induz a apoptose de células de glioblastoma (DING *et al*, 2011; DING *et al*, 2013). Em tumor de pulmão, Warth e colaboradores (2011) conseguiram identificar a superexpressão do transcrito e proteína AQP4 e associado com um prognóstico mais favorável.

No câncer de mama, a AQP1 foi correlacionada com mau prognóstico de pacientes com câncer de mama, e visto a existência de sua super expressão no citoplasma das células de câncer de mama. O estrogênio pode induzir a expressão de AQP1 ativando o elemento de resposta de estrogênio (ERE) no promotor do gene AQP1, resultando na formação de túbulos das células endoteliais vasculares (DAJANI, S; SARIPALLI, A; SHARMA-WALIA, N, 2018).

A AQP3 foi encontrada como um fator crítico para a migração de células de câncer de mama induzidas pelo fator de crescimento de fibroblastos-2 (FGF-2) no qual poderia aumentar a expressão de AQP3 e a migração celular através das vias de sinalização de FGFR-PI3K ou FGFR-ERK. Adicionalmente foi proposto que a AQP3 pode ser um fator limitante nos efeitos farmacológicos do 5'-desoxi-5-fluoropirimidina (5'-DFUR), usada na quimioterapia de tumores sólidos, uma vez que a exclusão do AQP3 reduz a eficácia do medicamento (DAJANI, S; *et al.*, 2018).

Em uma análise univariada, a superexpressão de AQP5 foi significativamente associada à sobrevida dos pacientes com câncer de mama precoce positivo para receptor de ER, PR e HER2. Além disso, uma análise de sobrevida multivariada revelou que a superexpressão de AQP5 pode ser um marcador de prognóstico independente de sobrevida para os pacientes com câncer de mama precoce (LEE, SJ; *et al.*, 2014).

Shi e colaboradores (2012), observaram que a expressão do mRNA da AQP4 estava maior em tecidos normais comparada ao tecido de câncer de mama.

Corroborando com os resultados de Shi., Zhu, *et al.* (2019), avaliaram a expressão de várias AQPs com o câncer de mama, dentre as quais foi vista que o mRNA da AQP4 está diminuído no câncer de mama, assim como a alta expressão desse mRNA foi associado a uma melhor sobrevida livre recidiva (*relapse-free survival*).

No caso de linhagens celulares de câncer de mama, a diminuição da expressão de AQP4 foi associada à redução na proliferação celular (LI; SUN; HAN, 2016).

O gene AQP4-AS1 (*Aquaporin 4 antisense RNA 1*) transcreve um RNA longo não codificante localizado no cromossomo 18q11.2 de função desconhecida (HALLADAY *et al.*, 2018). Entretanto, é conhecido que transcritos *antisense* podem inibir a expressão do transcrito *sense* através de modificações da cromatina e da terminação prematura da transcrição, e esse pode ser um mecanismo de ação da AQP4-AS1, embora ainda não tenha sido investigado.

Por meio de análises de bioinformática em dados de expressão gênica e estudos de associação com sobrevida, alterações de expressão gênica desse

lncRNA foi relacionada ao risco de câncer gástrico. Ademais, a expressão de *AQP4-AS1* estava intimamente relacionada à uma pior sobrevida global (XING, *et al.*, 2018).

Adicionalmente, a análise de dados de expressão do TCGA, mostram esse lncRNA desregulado no câncer de mama, porém, nem o SNP nem o lncRNA, onde este está mapeado, foram descritos com maiores detalhes.

3. OBJETIVOS

3.1 Objetivo Geral

— Verificar se há associação entre o câncer de mama e SNPs mapeados em RNAs longos não codificantes diferencialmente expressos nesse tumor.

3.2 Objetivos Específicos

— Identificar lncRNAs diferencialmente expressos em câncer de mama por meio de dados depositados no banco de dados TCGA.

— Identificar a presença de SNPs associados com câncer de mama através de dados depositados no catálogo GWAS, que estejam mapeados nas regiões dos lncRNA diferencialmente expressos em câncer de mama.

— Verificar se há associação entre o polimorfismo selecionado nos objetivos anteriores e a suscetibilidade ao câncer de mama em pacientes brasileiras.

— Investigar a associação do SNP com variáveis clínicas importantes nos pacientes com câncer de mama.

4. MATERIAL E MÉTODOS

Os experimentos foram realizados no laboratório de Citogenética Humana e Oncogenética (LabCHO), localizado na Universidade Federal do Paraná.

4.1 Análise de Bioinformática

Foram realizadas análises no banco de dados TCGA (THE CANCER GENOME ATLAS, 2018) em que identificamos lncRNAs que são diferencialmente expressos em amostras de câncer de mama em relação as amostras não tumorais.

Os dados de RNAs longos não codificantes foram obtidos na base de dados TANRIC (*The Atlas of non-coding RNA in Cancer*), que disponibiliza de forma gratuita os perfis de expressão de lncRNAs do TCGA (TANRIC, 2015). A partir desse banco de dados foram realizadas análises de expressão diferencial, utilizando o pacote em linguagem R Limma, entre 800 amostras tumorais de mama e 105 amostras normais de mama adjacente. Nesse pacote a expressão diferencial é avaliada através de um teste t Bayesiano utilizando um valor de $p < 0,001$, afim de identificar quais lncRNAs estão diferencialmente expressos nos tecidos tumorais.

Adicionalmente, avaliamos na lista de SNPs associados com câncer de mama, catalogados no GWAS CATALOG, as regiões mapeadas em lncRNAs evidenciados como diferencialmente expressos em nossas análises.

Após identificação dos lncRNAs, analisamos a diferença de expressão desses nos diferentes subtipos de câncer de mama através do portal TANRIC, também com os dados de expressão gênica depositados no TCGA.

Após essa análise, foi dada preferência aos SNPs que apresentavam os alelos menos frequentes em, no mínimo, 20% dos pacientes analisados em bancos de dados e também consideramos a amplitude de diferença de expressão nos subtipos moleculares de câncer de mama disponíveis na plataforma TANRIC.

4.2 Casuística

Para análise das frequências genotípicas do SNP selecionado, foram realizadas análises de amostras de pacientes com câncer de mama provenientes

do Hospital Nossa Senhora das Graças (HNSG) e amostras controles provenientes do banco do Laboratório de Imunogenética e Histocompatibilidade (LIGH) da UFPR. As amostras foram coletadas sob aprovação do comitê de ética em Pesquisa em Seres Humanos do Setor de Ciências da Saúde, UFPR, sob o número CAAE: 67029617.4.0000.0102. Todas as pacientes assinaram o termo de Consentimento Livre e Esclarecido.

Utilizando as frequências alélicas divulgados pelo GWAS, realizamos o cálculo de tamanho amostral, considerando o intervalo de confiança de 95% e a prevalência dos alelos menos frequentes em 20% da população, estimamos a genotipagem mínima de 300 amostras e 300 controles necessários para a produção de dados confiáveis para a conclusão deste trabalho (BEIGUELMAN, 1988).

Como controle foi utilizada uma amostra de 312 mulheres consideradas saudáveis do banco de doadores de medula óssea do Laboratório LIGH, com idade média de $47,66 \pm 4,69$.

Todas as amostras tumorais foram processadas no Laboratório de Citogenética Humana e Oncogenética do Departamento de Genética do Setor de Ciências Biológicas da UFPR. As amostras estão caracterizadas de acordo com características histopatológicas e imunoistoquímicas. As informações sobre os parâmetros clínicos e histopatológicos foram obtidas a partir dos laudos anatomopatológicos e o número de informações é variável entre as pacientes (TABELA 2)

TABELA 2. Dados Clínicos e Histopatológicos de pacientes com Câncer de mama.

Casos de Câncer de Mama n= 306					
Histologia	%	n	Grau Tumoral	%	N
Ductal	68%	209	I	7%	22
Lobular	10%	30	II	38%	115
Mucinoso	3%	8	III	19%	59
Misto ducto-lobular	6%	17	Sem informação	36%	110
Outros	9%	29			
Sem informação	4%	13			

Subtipo	%	n	Metástase nos	%	N
Imunoistoquímico			Linfonodos		
Luminal A	26%	79	Presença	28%	86
Luminal B	43%	132	Ausência	58%	176
HER2+	6%	17	Sem informação	14%	44
Triplo Negativo	9%	29			
Sem informação	16%	49			

FONTE: o Autor (2020).

4.3 Análise de polimorfismos

4.3.1 Extração de DNA

A extração de DNA das amostras tumorais, foi realizada de acordo com o protocolo padrão utilizado no Laboratório LabCHO. As amostras previamente reservadas a -80°C foram descongeladas e 30mg de amostras foram adicionados a 100ul de tampão proteinase K e 20ul de proteinase K (10ug/ml) em um tubo de 2ml. As amostras, posteriormente, foram incubadas a 56°C por cerca de 18 horas. Subsequentemente, adicionou-se 1 volume de fenol (aproximadamente 500µl) à reação, que foi homogeneizada por 15 segundos e centrifugada a 14000 rpm por 5

minutos. O sobrenadante retirado foi inserido em um novo tubo de 2ml, que passou pelo mesmo processo descrito anteriormente. Novamente, o sobrenadante foi retirado e adicionado a um novo tubo com adição de acetato de amônio 10M (100µl) e etanol absoluto (1ml) para a precipitação do DNA e adicionado os tubos no freezer até o dia seguinte.

No terceiro dia foi realizada a lavagem do DNA, iniciando com a centrifugação da solução a 14000 rpm por 15 minutos, desprezando-se o sobrenadante. Sucessivamente, foi adicionado 1ml de etanol 70% gelado, homogeneizando a amostra que foi levada novamente para centrifugação a 14000 rpm por 5 minutos. O sobrenadante foi, então, desprezado. Por fim, após a secagem dos tubos, o DNA foi ressuspendido em 50µl de água.

Para as amostras controle, foram adicionados junto a amostra: 240ul de água ultra pura, 80ul de tampão da proteinase K (5x), 40ul de proteinase K (10mg/ml) e 20ul de SDS 20%. Todas essas adições foram homogeneizadas e incubadas em banho maria a 56°C por 18 horas. Após, foi adicionado 100ul de solução saturada de NaCl (5M) e centrifugado a 10000 rpm por 2 minutos. O sobrenadante foi transferido para um novo tubo e centrifugado novamente a 10000 rpm por 2 minutos. Novamente, o sobrenadante foi transferido para um novo tubo e adicionamos, então, 1ml de etanol absoluto, invertendo o tubo gentilmente algumas vezes. Realizamos uma nova centrifugação a 10000rpm por 2 minutos e descartamos o sobrenadante. Repetimos o último passo com etanol 70% e, após a centrifugação, descartamos o sobrenadante. Ao fim, ressuspendemos o DNA em 40ul de água ultrapura e armazenamos a -20°C.

4.3.2 PCR alelo específica

Após a extração de DNA, as regiões específicas do polimorfismo foram amplificadas a partir de PCRs, utilizando primers de sequência-específica, capazes de diferenciar cada alelo. Os primers utilizados para amplificar as regiões de interesse nessa pesquisa estão descritos na tabela 3.

TABELA 3. Primers utilizados para realização de PCRs alelo específicas na genotipagem das amostras de pacientes com câncer de mama e amostras controles.

LncRNA	FORWARD 5' - 3'	REVERSE 5' - 3'	TAMANHO
AQP4-AS1- Alelo C	AGCTCCAGTGCTATTT G	ACAGGTCAAGGAAATGC	167pb
AQP4-AS1- Alelo G	GTTGTAGAAGGCACAGTTG	AGGACAAGTCTAACTAGG G	117pb

FONTE: o Autor (2020).

Em condições adequadas de PCR, os primers se ligam à região específica onde se encontra o SNP a ser observado (FIGURA 5), favorecendo sua amplificação. Para esses conjuntos de primers as condições ideais estão descritas na tabela 4.

FIGURA 5. Sequência Fasta da fita antisense e primers do lncRNA *AQP4-AS1*.

```

CTCTGACAAA GGAGAAGGAA AATGGGGAGG TTAAGGACCC TACACCTAAA CCTTTAGATC
ATTCTCCAAA GCCTGGATCT CTGGAGGTGA GCTGCATTAA GCTGACTTGA ACTGAATCAG
ACCATACCAG GCCAGTCAGT AATACTGCAT CACCTGGAGG GCCTCACCCA ACACTCTCCT
CTGGGAAAAC GATGAGAATG GTGATGTAAA ATGCTAGAAA GTGAATTAAC TGCTCCCCAA
ACTTCATGCT ATTTTAAACAT TCTGTGGTCA TGTGTTTTTT TGTTTTTTGG TAAACTCTGA
CAAATGCTTA GAATTTTGTC AAAACAGACT GAACTGCAAT CATGCTGCGT TACACTGACA
TTATTACACAC TTGCTAATGA TTAACCTCTCT TTGACACTGT AGCGTTGTAG AAGGCACAGT
TGAACAAACA GCTTCAGGGG GGCCAGAAAT GAGTTACACG AGACTGAGCC AACTGACTTG
GAGGAGCTCC AGTGCTATTT
S C>G
CCTAGTTTAG ACTTGTCCTA AGGGGCATTT CCATTTGGGG ACCTGAAAGTC CTGGAATTCC
TTCCTCTTCC TGGTTCAACA AAACCTCTAAT TTATTGACCT GCAGGGCACA GTCTAAGACA
TCTGTAAAGT TTGGCATTTC CTTGACCTGT ACAGAGAATT CTGTTGAGTA AAAATGGTTT
GCCTTTTTCA TTCCAGTTTT TGTAGTGTGT ATGCATCCCG ACTCTGATCT AAAGGAAGTA
CCTTCGTATA AGCCCAAGAA ACAAAAGGGT TTTGCTTCTT TTGCATCAAA GCTGCTCCTT
TTGGAGCATC ACCTCAATGC AGTCTTCTCT TTTCAGTGTG ACAAAACCTT GGATAAATAT
CCAAGAATCC TTCCTTTACT ACACAAGGAG AAAAGGAGAA AGTGTATTGA ACAGCATTCT
TTGTGCTTGG CATGAAGAAA ATCAGTCCTG CAAGAATGTG AGAAAAATTTT GGTATATAAG

```

LEGENDA: Em verde primers para amplificação do alelo G, em amarelo primers para amplificação do alelo C. FONTE: NCBI: NC_000018.10 (2020).

TABELA 4- Ciclagem da reação do lncRNA *AQP4-AS1*.

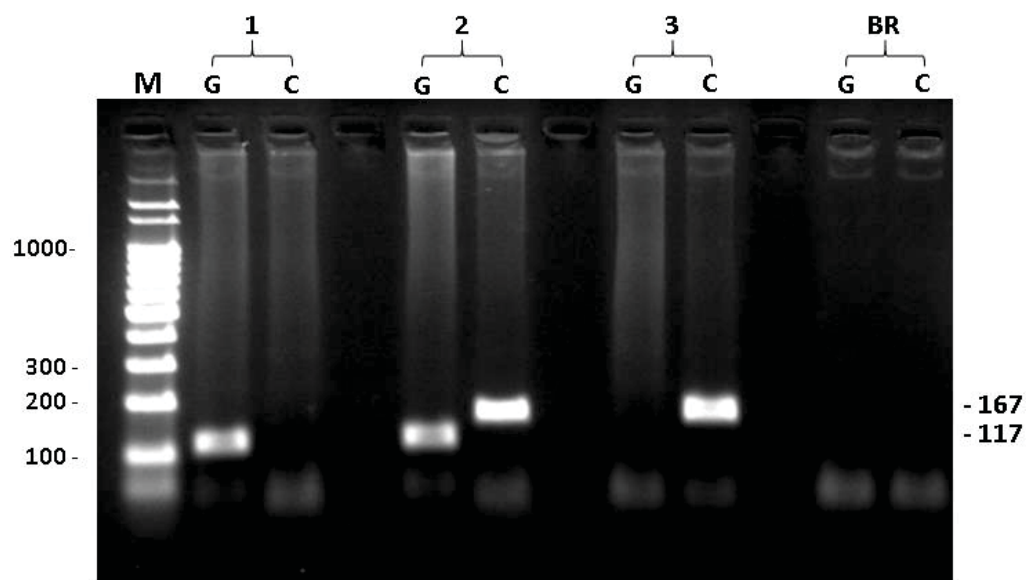
Etapa do Ciclo		Temperatura	Tempo
Desnaturação inicial		95°C	10min
35x	Desnaturação	96°C	30seg
	Anelamento dos iniciadores	60°C	30seg
	Alongamento	72°C	30seg
	Alongamento final	72°C	10min

FONTE: o Autor (2020).

As PCRs foram realizadas a partir de 2µl de DNA em concentração de 20ng/µl e 160 pmol de primer específico na presença de Master Mix para PCR convencional (1x) desenvolvido pelo IBMP, ICC/FioCruz.

Os resultados foram interpretados por meio de análise de eletroforese em gel de agarose 2% corado com Gel Red (Biotium), como demonstrados na figura 6.

FIGURA 6. Padrão eletroforético dos fragmentos amplificados.



LEGENDA. M- Marcador de peso molecular; 1- Amostra homozigota GG; 2- Amostra heterozigota CG; 3- Amostra homozigota CC; BR- brancos. Direita: Tamanho esperado do fragmento. FONTE: o Autor (2020).

4.3.3 Sequenciamento pelo Método Sanger

O método é baseado na incorporação de desoxinucleotídeos (dNTPs) e de didesoxinucleotídeos terminadores (ddNTPs) a uma cadeia de DNA em crescimento, tendo como molde o DNA de interesse (SANGER *et al*, 1977). Para confirmação da especificidade das amplificações, a região esperada foi sequenciada.

Os *primers* (TABELA 5) utilizados para a PCR deste fragmento estavam concentrados em 25µmoles.

TABELA 5- Primers utilizados no Sequenciamento do método Sanger.

LNC RNA	FOWARD	REVERSE
<i>AQP4-AS1</i>	GTTGTAGAAGGCACAGTTG	ACAGGTCAAGGAAATGC

FONTE: o Autor (2020).

Após a realização da PCR as amostras foram submetidas ao Método de Purificação de Produtos de PCR que utiliza Exonuclease I (Exo I) para eliminar o excesso de *primers* da reação e *Shrimp Alkaline Phosphatase* (SAP) para degradar excesso de nucleotídeos provenientes da PCR.

Para cada amostra foram adicionados 0,2µl de Exonuclease I (1 unidade/µl), 0,6µl de SAP (0,66 unidades/µl) e 3µl do produto de PCR. Esta mistura foi incubada a 37°C por 45 minutos em Termociclador *Mastercycler gradiente* (Eppendorf) e posteriormente a 80°C, por 10 minutos, também em termociclador para inativação das enzimas.

Em seguida as amostras foram submetidas à reação de sequenciamento, utilizando o kit *BigDye Terminator v3.1 Cycle Sequencing* (Applied Biosystems). Para cada reação de sequenciamento, foram utilizados 0,5µl de *Ready Reaction Premix* (2,5X), 0,2 µl do primer foward (25µmoles), 2µl de Tampão *BigDye Sequencing Buffer* (5X), 1µl do produto de PCR tratado pelo EXO/SAP e água ultrapura até completar 10µl. O programa da ciclagem desta etapa está representado na tabela 6.

TABELA 6- Ciclagem da reação de sequenciamento realizada no aparelho *MasterCycler Gradient*, *Eppendorf*.

Etapa do Ciclo		Temperatura	Tempo
Desnaturação inicial		96°C	1min
30x	Desnaturação	96°C	30seg
	Anelamento dos iniciadores	52°C	30seg
	Alongamento	60°C	2Min

FONTE: o Autor (2020).

Para retirada dos *primers*, terminadores fluorescentes e outras impurezas foi realizada a purificação das reações por meio da adição 2 µl de acetato de sódio 3M (7,5 pH) e 25 µl de etanol absoluto 100%, centrifugado por 40 minutos a 14000 rpm e desprezado o sobrenadante. O etanol absoluto atua depletando a camada de hidratação dos ácidos nucleicos o que expõe as cargas negativas dos grupamentos fosfato. O acetato atua neutralizando estas cargas expostas o que permite a formação de precipitados de DNA. Na sequência foram adicionados 50 µl etanol 80%, centrifugação por 10 minutos a 14000 rpm, em seguida foi desprezado o sobrenadante e colocado para evaporação do etanol no termociclador com a tampa aberta a 60°C durante 5 minutos.

O produto precipitado foi analisado em sequenciador automático ABI-3530 Applied Biosystems, do Departamento de Bioquímica da Universidade Federal do Paraná e os dados coletados pelos softwares ABI-3530 Collection e Sequencing Analysis.

4.4 Análise estatística

Para os testes de frequências genotípicas do grupo controle e do grupo de pacientes, foi utilizado o teste de desvios das proporções do teorema de Hardy-Weinberg por Qui-quadrado. Adicionalmente, por ser um teste particularmente indicado para estudos de caso-controle, foi utilizado o cálculo de *Odds ratio* (OR), e

também o teste de Qui-Quadrado para avaliar se as variáveis (câncer de mama e SNPs) são ou não independentes.

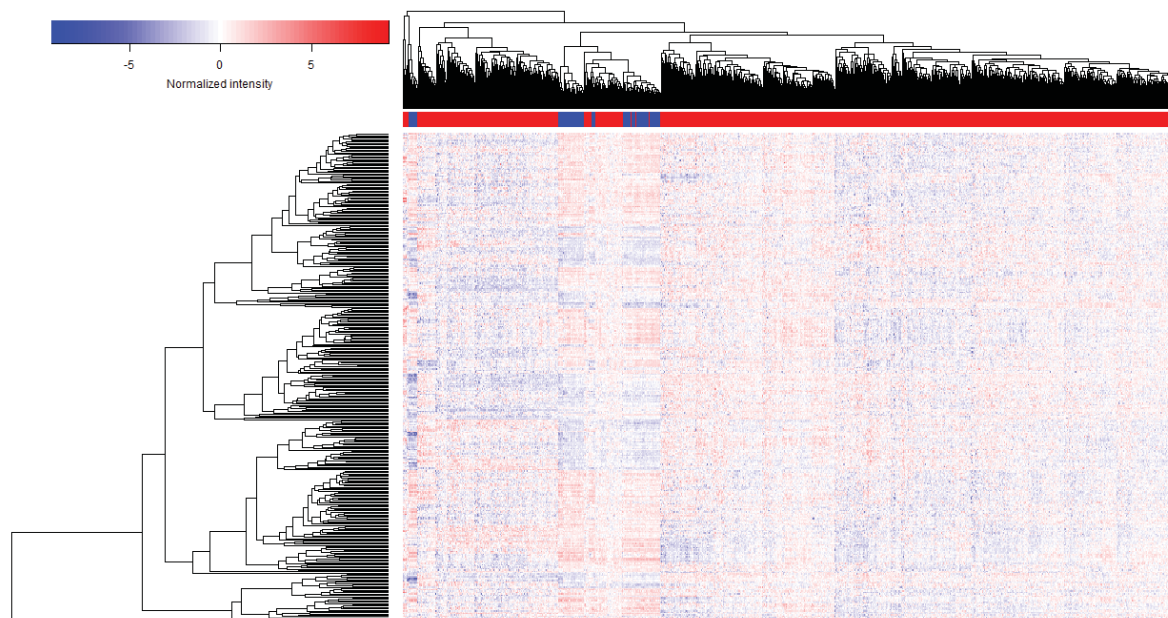
Os testes estatísticos foram realizados com o *software* R (R CORE TEAM, 2017) com auxílio dos pacotes Nortest e readxl (GROSS; LIGGES, 2015; WICKHAM; BRYAN, 2017). Para todos os testes acima descritos foram considerados como significativos valores de $p < 0,05$.

5. RESULTADOS

5.1 LncRNAs Diferencialmente Expressos em Câncer

A partir da análise de dados brutos de 12.000 lncRNAs (TCGA), obtivemos 1334 lncRNAs diferencialmente expressos em câncer de mama comparados a tecido normal com $p < 0,001$, e *fold change* de $\pm 1,5$ ($= \log_{FC} 0,6$), representados na figura 7.

FIGURA 7. *HeatMap* de lncRNA diferencialmente expressos em câncer de mama. A barra de cores abaixo do dendrograma representa amostras tumorais (vermelho) e amostras não tumorais (azuis).



FONTE: o Autor (2020).

5.2 LncRNAs com SNPs associados com Câncer de Mama

Após a seleção dos lncRNAs diferencialmente expressos, realizamos o cruzamento desses dados com a tabela de SNPs evidenciados em um estudo GWA. Nesse estudo de 137.045 casos de câncer de mama e 119.079 controles analisados, foram encontrados 38.134 SNPs associados ao risco de câncer de mama (MICHAILIDOU *et al.*, 2017).

Cruzando os dados dos lncRNAs diferencialmente expressos e dos SNPs associados com câncer de mama, identificamos 22 lncRNAs. Adicionalmente, observamos a diferença de expressão destes nos subtipos de câncer de mama (TANRIC), assim como, aqueles que apresentavam uma frequência do alelo mais raro (*minor allele frequency* – MAF) $> 0,2$, uma vez que isso facilitaria a análise em uma coorte menor, com maior chance de encontrar os alelos menos frequentes.

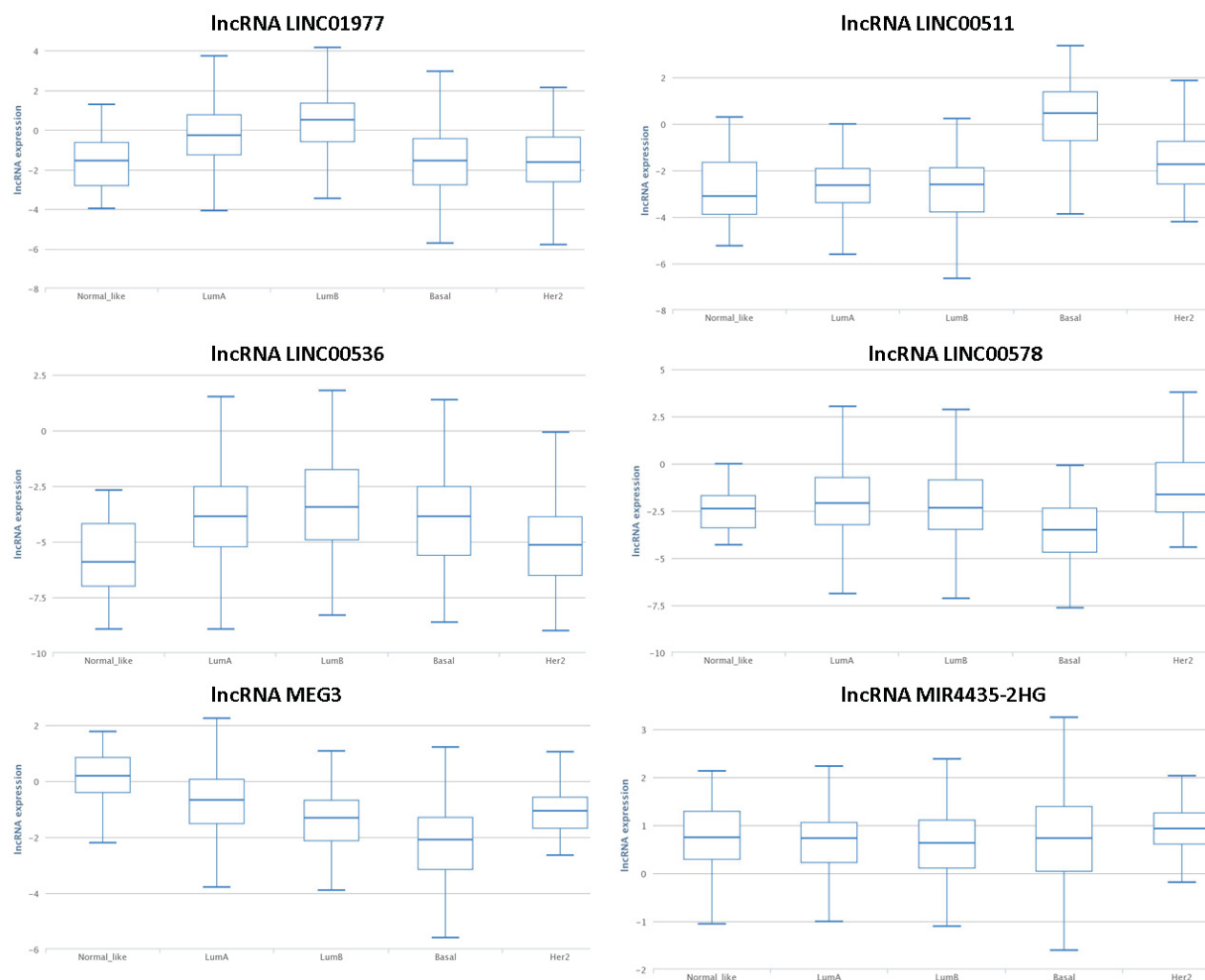
Dos 22, ficamos com 14 lncRNAs que também apresentavam grande diferença de expressão nos subtipos moleculares de câncer de mama (FIGURA 8 e 9).

FIGURA 8. Diferença de expressão dos lncRNA nos subtipos moleculares de câncer de mama. Destacado em vermelho o lncRNA analisado neste trabalho.



LEGENDA: Da esquerda para a direita os dados referem-se a *Normal-like*, Luminal A, Luminal B, *Basal-like*, Her2+. FONTE: *The Atlas of ncRNA in Cancer- Tanric* (2020).

FIGURA 9. Diferença de expressão dos lncRNA nos subtipos moleculares de câncer de mama.



LEGENDA: Da esquerda para a direita os dados referem-se a *Normal-like*, Luminal A, Luminal B, *Basal-like*, Her2+. FONTE: *The Atlas of ncRNA in Cancer- Tanric* (2020).

A partir de estudos das informações presentes na literatura sobre a possível importância desses lncRNAs, nos chamou atenção o SNP mapeado no lncRNA *AQP4-AS1*, uma vez que não existe nenhum trabalho na literatura evidenciando esse lncRNA em câncer de mama e trabalhos recentes sugerem a importância da proteína AQP4, produzido na fita complementar a esse lncRNA e que poderia ser regulada por esse lncRNA sob influência desse SNP. Portanto, escolhemos o

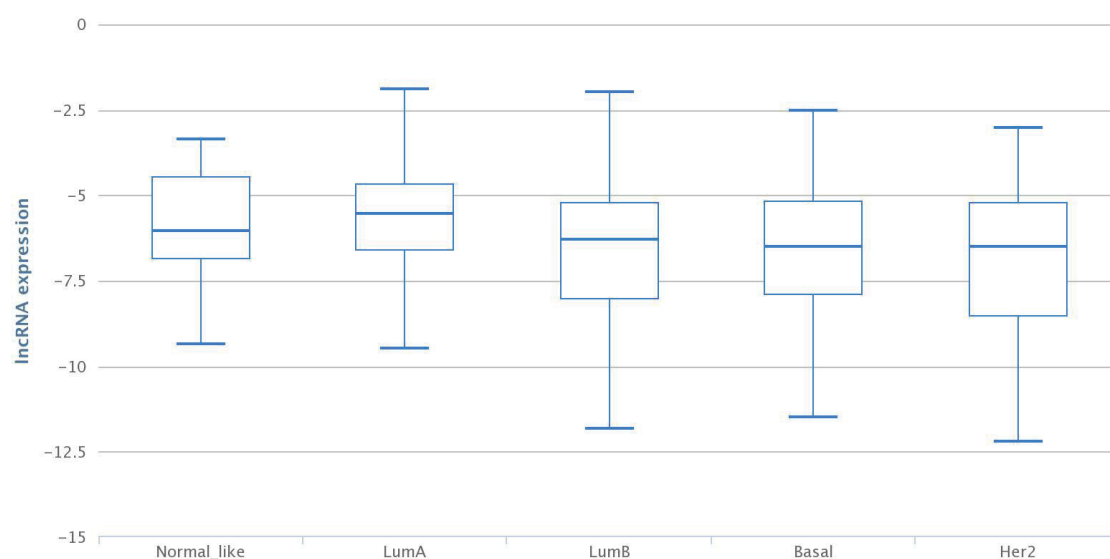
rs527616, mapeado no lncRNA *AQP4-AS1* para avaliar sua frequência em uma segunda população de amostras de Curitiba-PR. (TABELA 7).

TABELA 7- LncRNAs selecionados e seus respectivos dados, em vermelho o LncRNA selecionado para este estudo. P-value¹ referente ao logFC, P-value² associação do LncRNA ao câncer de mama.

LNCRNA	LOGFC	P-value ¹	TANRIC	GWAS	OR	P-value ²	IC	MAF
AC020916.1	-1,99	1,27E-147	<ER+	rs2594714	1,04	1 x 10-8	1.01-1.05	0,42
AC093297.2	1.86	2.70E-15	<ER+>HER+	rs7716600	1,24	7 x 10-7	1.14-1.34	0,27
AL358075.2	1.45	3.39E-11	<HER2+	rs1707302	1,04	3 x 10-8	1.02-1.05	0,37
AQP4-AS1	0,87	9,37E-10	>ER+	rs527616	1,03	7 x10-15	1.02-1.05	0,23
CASC16	1,92	9,57E-15	>ER+	rs4784227/rs3803662(TAG)	1,23	4 x10-117	1.2-1.25	0,25
CRYZL2P-SEC16B-201	-0.80	1.90E-12	>HER2+	rs575908	1,03	3 x 10-6	0.019-0.045	0,35
LINC-PINT	-0,69	1,14E-06	>HER-	rs68056147	1,05	5 x10-13	0.037-0.064	0,24
LINC02224	3.92	3.14E-100	<ER+>HER+	rs4415084	1,17	8 x 10-11	1.11-1.22	0,46
LINC01977	1.48	5.16E-19	<ER+	rs745570	1,03	4 x 10-10	1.01-1.05	0,38
LINC00511	1.72	2.70E-14	>ER+	rs11652463	1,04	8 x 10-8	0.025-0.055	0,43
LINC00536	1,05	1,43E-05	>ER+	rs13267382	1,03	2 x10-11	1.03-1.07	0,34
LINC00578	1.91	2.70E-16	<ER+>HER+	rs7430456	1,02	5 x 10-6	0.017-0.042	0,44
MEG3	-1,93	2,53E-47	>ER+	rs2295389	1,03	2 x10-6	0.02-0.049	0,27
MIR4435-2HG	0.88	6.77E-10	<HER2+	rs200484318	1,04	7 x 10-6	0.024-0.062	0,36

FONTE: o Autor (2020).

FIGURA 10. Expressão do *AQP4-AS1* em diferentes subtipos de câncer de mama.



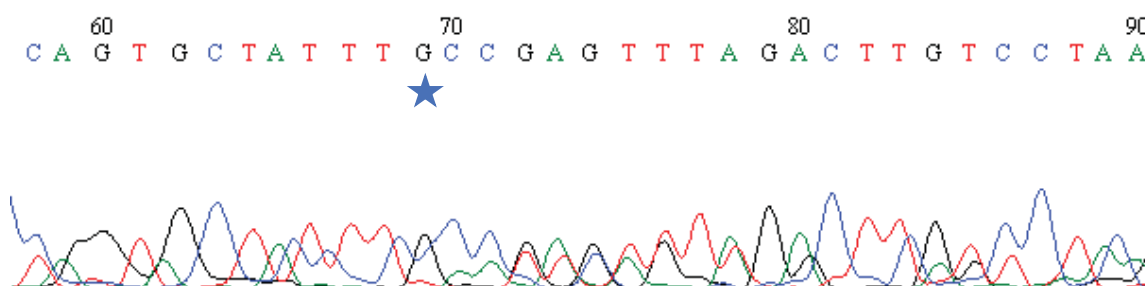
FONTE: *The Atlas of ncRNA in Cancer- Tanric* (2020).

5.3 Genotipagem dos SNPs

Após a seleção do lncRNA *AQP4-AS1* e seu SNP, realizamos o desenho dos *primers* para genotipagem das amostras e a padronização das condições ideais de PCR para amplificação da região desejada.

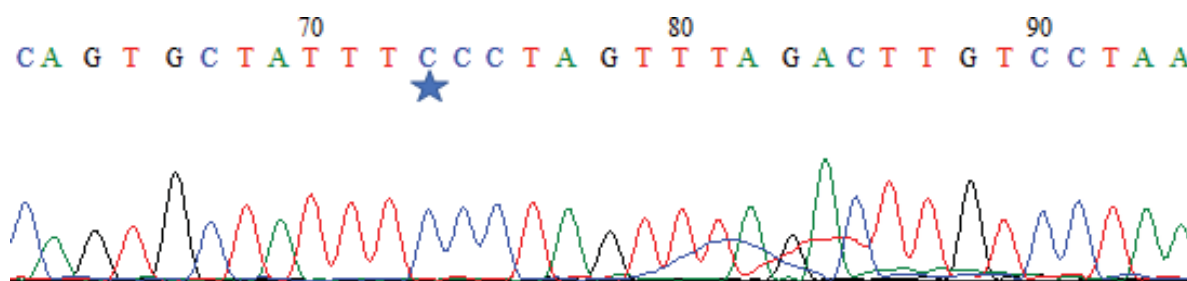
Para confirmação da exatidão dos resultados das PCRs, foi realizado o sequenciamento da região que foi analisada (FIGURAS 11 e 12).

FIGURA 11. Sequenciamento do lncRNA *AQP4-AS1*, local do SNP marcado pela estrela, amostra de genótipo GG.



FONTE: o Autor (2020).

FIGURA 12. Sequenciamento do lncRNA *AQP4-AS1*, local do SNP marcado pela estrela, amostra de genótipo CC.



FONTE: o Autor (2020).

5.4 Frequências Alélicas e Genotípicas

As genotipagens do lncRNA *AQP4-AS1* foram realizadas em 306 pacientes e 312 controles. As frequências genotípicas observadas em pacientes e controles foram respectivamente 3% (n=9) e 8% (n= 25) homozigotos C/C; 55% (n= 167) e 41% (n=129) heterozigotos C/G; 42% (n= 130) e 51% (n=158) homozigotos G/G. (TABELA 8). A frequência alélica do SNP rs527616 calculada no grupo de pacientes foi de C: 30% e G: 70% e no grupo de controles C: 29% e G: 71%, muito semelhante as frequências alélicas em outros estudos divulgados em diferentes bancos de dados (TABELA 9).

As frequências genotípicas observadas para este gene no grupo controle mostraram-se em equilíbrio de Hardy-Weinberg, sugerindo que a amostra controle analisada foi retirada de população em equilíbrio ($\chi^2 = 0,068$; $p > 0,70$). As frequências genotípicas observadas para este gene nas pacientes não estão em equilíbrio de Hardy-Weinberg, indicando que a amostra analisada foi retirada de população que não se apresenta em equilíbrio ($\chi^2 = 26,65$; $p < 0,001$).

TABELA 8. Frequências genotípicas e alélicas dos pacientes e controles.

GENÓTIPO/ ALELO	PACIENTES (n=306)		CONTROLES (n=312)	
	%	n	%	n
CC	3%	9	8%	25
CG	55%	167	41%	129
GG	42%	130	51%	158
ALELO C	30%		29%	
ALELO G	70%		71%	
MAF	0,30		0,29	

FONTE: o Autor (2020).

TABELA 9- Frequência do SNP rs527616 em diferentes estudos.

ESTUDO	TAMANHO DA AMOSTRA	ALELO DE REFERÊNCIA	ALELO ALTERNATIVO
1000GENOMES	1189 / 5008	C: 0,24	G: 0,77
TOPMED	34949 / 125568	C: 0,28	G: 0,73
GNOMAD	9920 / 31334	C: 0,32	G: 0,69
ESTONIAN	2060 / 4480	C:0,46	G: 0,54
PAGE_STUDY	15281 / 78698	C:0,19	G: 0,81
ALSPAC	1429 / 3854	C: 0,37	G: 0,63

FONTE: *National Center for Biotechnology Information - NCBI databases* (2020).

5.5 Estudo de Associação

A comparação das frequências genotípicas entre pacientes e controles apresentou associação com a suscetibilidade com o câncer de mama quando comparados os genótipos GG vs GG (OR = 1,38901; p= 0,035; IC = 95% 1,011 - 1,907) e CG vs CG (OR = 1,704; p= 0,0009; IC = 95% 1,2395 - 2,3435), ao comparar os genótipos CC vs CC foi observado um valor de OR =0,3478; p= 0,004; IC = 95% 0,1596 - 0,7581, conferindo fator protetivo ao desenvolvimento do câncer de mama.

Adicionalmente realizamos o teste de *Odds ratio* para o modelo recessivo (GG/CG vs CC), e observamos que o resultado corrobora com as análises de OR entre os genótipos CC e CG estando associados com a suscetibilidade com o câncer de mama OR = 2,874; p= 0,004; IC = 95% 1,319 - 6,264.

Para verificar a existência de relação alélica, realizamos a comparação entre os alelos C vs G, que não apresentou relação estatística significativa, OR = 0,92; p= 0,535; IC = 95% 0,7269 - 1,1857. TABELA 10.

TABELA 10. Análises estatísticas de *Odds Ratio*.

<i>Odds Ratio</i> (IC 95%)				
C vs G	GG vs GG	CG vs CG	CC vs CC	GG/CG vs CC
0,92 (0,72-1,18)	1,38 (1,01-1,90)	1,70 (1,23-2,34)	0,34 (0,15-0,75)	2,87 (1,31-6,26)
p= 0,535	p= 0,035*	p= 0,0009*	p= 0,004*	p= 0,004*

FONTE: o Autor (2020). * <0.05

5.6 Análise dos Subtipos e Parâmetros Histopatológicos

Utilizamos o modelo dominante para avaliar a correlação entre os genótipos e os subtipos luminais (A e B) e triplo negativo, assim como foram correlacionados aos parâmetros histopatológicos (metástase nos linfonodos regionais e grau de diferenciação, selecionados de acordo com a disponibilidade das informações presentes nos laudos das pacientes.

O Qui-Quadrado de homogeneidade não demonstrou diferenças significativas quando o total de pacientes com genótipo considerado de risco foi comparado com aqueles sem genótipo de risco, para os subtipos $\chi^2 = 2,17$ e $p > 0,14$ e nenhum dos parâmetros histopatológicos: invasão de linfonodos regionais $\chi^2 = 1,92$ e $p > 0,16$ (GL=1); grau de diferenciação tumoral $\chi^2 = 0,761$ e $p > 0,65$ (TABELAS 11 e 12).

TABELA 11. Análise de Qui-Quadrado de homogeneidade para subtipos imunoistoquímicos.

	Luminais	Triplo Negativo
GG/CG	148	27
CC	3	2
Qui-Quadrado	$\chi^2 = 2,17$ p>0,14	

FONTE: o Autor (2020).

TABELA 12. Análise de Qui-Quadrado de homogeneidade para metástase nos linfonodos e grau de diferenciação tumoral.

	Linfonodo Positivo	Linfonodo Negativo	Grau I	Grau II	Grau III
GG/CG	82	173	21	112	56
CC	4	3	1	3	3
Qui-Quadrado	$\chi^2 = 1,92$ p=0,16		$\chi^2 = 0,761$ p=0,65		

FONTE: o Autor (2020).

6. DISCUSSÃO

O câncer de mama é a neoplasia mais comum em mulheres em todo o mundo. De acordo com as últimas estatísticas mundiais do Globocan 2018, foram estimados 2,1 milhões de casos novos de câncer e 627 mil óbitos pela doença. Embora ocorra em todo o mundo, sua incidência e mortalidade podem variar consideravelmente pela influência de muitos fatores complexos, como estrutura populacional, estilo de vida, ancestralidade, ambiente, *status* socioeconômico, acesso a cuidados de alta qualidade e fatores genéticos (HORTOBAGYI *et al.*, 2005).

Entre os fatores envolvidos na tumorigênese, incluindo do câncer de mama, estão os lncRNAs (LIU, SHARMA, WATABE, 2017). Além dos fatores de risco epidemiológicos, evidências crescentes sugerem que os polimorfismos de nucleotídeo único (SNPs) podem ser de fundamental importância na suscetibilidade genética ao câncer de mama (LI *et al.*, 2018).

Estudos GWA, disponibilizados pelo GWAS CATALOG (2018), indicam que mais de 1700 associações entre SNPs e câncer de mama já foram avaliadas, das quais muitas foram observadas nas regiões de ncRNAs. Evidências cumulativas indicam a ligação entre SNPs relacionados a ncRNA e várias doenças humanas. No entanto, o conhecimento dos SNPs relacionados aos ncRNAs e seus potenciais mecanismos funcionais em doenças humanas precisam ser mais estudados (NING, S; ZHANG, Y; 2018.).

Estudos confirmaram que os SNPs em lncRNAs podem estar envolvidos na tumorigênese, em mecanismos secundários dos RNAs mensageiros correspondentes ou alterando seus parceiros de interação. Isso significa que os locais de suscetibilidade funcional desempenham um papel fundamental na ocorrência e progressão do câncer. (LI, J; *et al.*, 2018).

Como exemplo temos o lncRNA HOTAIR (*HOX transcript antisense intergenic RNA*), um dos lncRNAs mais bem descritos devido sua associação com vários tipos de cânceres, no qual a presença do rs920788 foi descrito associado com a mudança da expressão desse lncRNA e a suscetibilidade a cânceres (LI, J; *et al.*, 2018).

Através de análises de expressão e de estudos GWA é possível observar a associação de lncRNAs em câncer de mama e vários outros tipos de cânceres, reforçando sua importância no contexto celular e seu potencial como biomarcadores.

A nossa análise em bancos de dados, indicou que dos 1334 lncRNAs como diferencialmente expressos em câncer de mama, 22 possuem SNPs evidenciados no GWAS. Desses, 14 também possuem uma alta frequência de variação (MAF>0.2) e diferença de expressão entre os subtipos de câncer de mama.

Alguns dos lncRNA apresentados em nossa análise já foram estudados em diferentes tipos de cânceres, como o CASC16 que apesar de pouco conhecido, há indicações de que a presença do rs4784227 diminui a expressão do gene *TOX3* (COWPER-SAL·LARI *et al.*, 2012), que funciona como um supressor de tumor. Sua baixa expressão está associada na sobrevida de pacientes com o subtipo luminal A (GUDMUNDSDOTTIR *et al.*, 2012).

O lncRNA LINC-PINT foi descrito em diferentes tipos de cânceres e atuando como um supressor de tumor na redução do fenótipo invasivo de células cancerígenas (BÉJAR, *et al.*, 2017), além de ter sido relacionado com o câncer de pâncreas e câncer gástrico (LI, *et al.*, 2016; FENG, *et al.*, 2019.)

Outro lncRNA dentre os 14 de nossa lista, foi o LINC00511, que foi descrito participando no desenvolvimento de osteossarcoma (YAN, *et al.*, 2018), além de ter sua expressão significativamente aumentada em células de câncer hepático (WANG, *et al.*, 2019.). Em câncer de mama, foi demonstrado que o nível de expressão do LINC00511 estava elevado nos tecidos de células com subtipo triplo negativo comparadas com tecidos normais. A regulação negativa de LINC00511 diminuiu o crescimento e a invasão de células triplo negativas em comparação com amostras controles (LIU, *et al.*, 2019).

Li e colaboradores (2019), apresentaram dados que demonstraram que a expressão de outro lncRNA em nossa lista, o LINC00536 está aumentado nos tecidos de câncer de bexiga comparado com controles, com menor sobrevida dos pacientes. Já análises do lncRNA LINC00578 indicaram que pacientes com alta expressão do mesmo tiveram uma taxa de sobrevida significativamente menor no câncer pancreático (ZHANG, *et al.*, 2018).

O lncRNA *MEG3* é o mais bem descrito e conhecido dentre os 14, vários pesquisadores já demonstraram sua importância como supressor de tumor em diversos tipos de cânceres e como participante na alteração dos níveis de expressão da p53 (ZHOU *et al.*, 2007). Cânceres como o colorretal (ZHU *et al.*, 2018), de pulmão (LU *et al.*, 2013), cervical (QIN *et al.*, 2013), gástrico (SUN *et al.*, 2013), de fígado (HE *et al.*, 2017), de ovário (SHENG *et al.*, 2014) e osteossarcomas (TIAN *et al.*, 2015) são outros exemplos em que foi observado a diminuição da expressão do *MEG3* nas células de câncer comparadas aos tecidos normais.

Estudos indicam que o silenciamento do lncRNA MIR4435-2HG inibiu a proliferação e metástase de células de câncer de pulmão *in vitro* e *in vivo* (QIAN, *et al.*, 2018). Por outro lado, Ouyang, *et al.*, (2019) demonstraram a presença de níveis elevados de MIR4435-2HG em tecido de câncer colorretal, onde esse aumento também estava relacionado com o estágio TNM e pior sobrevidas livre e global.

Dentre os 14 lncRNAs selecionados em nossas análises, alguns ainda não foram estudados (AC020916.1, AC093297.2, AL358075.2, CRYZL2P-SEC16B, LINC02224, LINC01977), demonstrando que ainda são necessárias várias pesquisas para analisar as participações desses lncRNAs com diversos tipos de câncer, incluindo o de mama.

Analisando a importância já descrita de vários lncRNAs, nosso levantamento de dados disponibiliza informações importantes para futuros trabalhos, principalmente para aqueles lncRNAs que ainda não foram estudados.

Dos 14 lncRNAs destacados, escolhemos o SNP rs527616, localizado no lncRNA *AQP4-AS1* para estudo na população brasileira.

Dados de frequências alélicas divulgados por bancos de dados como *1000 Genomes*, *gnomAD – Genomes* e *TOPMED* (Tabela 9), demonstram que a frequência do alelo menos frequente (C), varia em diferentes populações, uma vez que, diferentes populações com diferentes ancestralidades e ambientes podem apresentar variações alélicas (UPADHYAYA et al., 2016).

Considerando os resultados de nossas análises, podemos observar que existe uma variação do MAF tanto em nossas amostras de pacientes (0,3) como em controles (0,29), se comparadas aos dados divulgados de outras populações, poderemos visualizar que os resultados dos estudos *gnomAD – Genomes* (0,31) e *TOPMED* (0,27) possuem o MAF muito semelhantes e seus dados são compostos em geral, de amostras de populações europeias. Portanto, considerando que as populações do Sul do Brasil possuem uma considerável ascendência europeia, os resultados comparativos observados já eram esperados (PROBST et al., 2000).

Ao analisarmos as frequências genotípicas do grupo controle, observamos que estas estão em concordância com as frequências esperadas pelo teorema de Hardy-Weinberg, confirmando que as amostras controles foram retiradas de uma população em equilíbrio, diferentemente das frequências genotípicas observadas no grupo de pacientes, que não se distribuíram de acordo com o equilíbrio de Hardy-Weinberg ($\chi^2 = 26,65$; $p < 0,001$).

Embora as frequências alélicas para ambos os grupos tenham sido semelhantes, observou-se na amostra de pacientes com câncer de mama uma

frequência acima do esperado para heterozigotas CG e abaixo do esperado para homozigotas CC, fato que nos permitiu admitir a hipótese de que a presença do genótipo CC pudesse estar de alguma forma associada a diminuição a suscetibilidade ao câncer de mama.

Assim, analisamos as OR, comparando os diferentes genótipos (CC, CG e GG) entre pacientes e controles. Observamos que os genótipos GG (OR= 1,38, IC = 95% 1,011 - 1,907) e CG (OR=1,70, IC = 95% 1,2395 - 2,3435) sugerem associação ao aumento de risco ao câncer de mama. Adicionalmente realizamos o teste de OR para GG/CG vs CC, onde um aumento significativo do OR (2,87, IC = 95% 1,319 - 6,264) foi observado, corroborando com a hipótese de que os genótipos GG e CG participam com o aumento do risco de câncer de mama.

Para verificar se essa associação devida ao alelo ou ao genótipo, realizamos a análise isolada dos alelos C vs G entre pacientes e controles, onde não encontramos associação alélica (OR = 0,92, IC = 95% 0,7269 - 1,1857), sugerindo que o câncer de mama esteja associado ao genótipo e não a um alelo específico.

Concordando com nossos resultados, os estudos GWA indicam a presença do Alelo G do SNP rs527616 como o alelo de risco ao câncer de mama, como anteriormente citado, variações podem ocorrer em diferentes populações, resultando em diferentes interações que podem determinar que, alelos considerados de risco em uma população, podem ser considerados alelos de proteção em outras (UPADHYAYA et al., 2016). Em concordância com os dados divulgados por estudos GWA, calculamos o OR para o genótipo CC (OR = 0,34, IC = 95% 0,1596 - 0,7581) e verificamos que a presença desse genótipo sugere uma associação protetiva ao desenvolvimento com o câncer de mama.

Também, analisamos as frequências genotípicas em função de características clínico-patológicas os subtipos imunoistoquímicos luminais (A e B) e triplo negativo ($\chi^2 = 2,17$ $p > 0,14$), para metástase nos linfonodos ($\chi^2 = 1,92$ $p > 0,16$) e para grau de diferenciação tumoral ($\chi^2 = 0,761$ $p > 0,65$). Em nenhuma das comparações as diferenças foram estatisticamente significantes. Apesar dos resultados não terem sido significativos, mais estudos com amostras mais

consistentes são necessários para avaliar os subtipos de câncer de mama com diferentes comportamentos de agressividade e o rs527616.

Apesar do lncRNA *AQP4-AS1* não possuir função conhecida (HALLADAY *et al.*, 2018), estudos GWA indicam que a presença do rs527616 está associada a suscetibilidade ao câncer de mama.

De acordo com os dados de expressão disponibilizados pelo banco de dados TANRIC, observa-se uma diminuição na expressão do *AQP4-AS1* nos diferentes subtipos de câncer de mama. Ao analisarmos nossos dados de genotipagem, observamos que a presença do genótipo CC, fator protetivo, está diminuída em comparação ao genótipo heterozigoto. Assim, sugerimos que exista interação entre a diminuição do genótipo protetivo CC e a diferença de expressão do *AQP4-AS1* nos diferentes subtipos de câncer de mama.

A sugestão acima apresentada torna-se mais viável uma vez que é conhecido que SNPs podem interferir na expressão de um gene ao mudar a estrutura de um lncRNA, seu sítio de ligação a proteínas e mecanismos secundários dos RNAs mensageiros correspondentes ou, ainda, alterando seus parceiros de interação. (LI, J; *et al.*, 2018).

Corroborando com essas informações, sabe-se que genes *antisense* podem alterar a expressão de genes *sense* de diversas formas, como metilação do DNA, modificação da cromatina, variação de isoformas e alteração de instabilidade de RNA (PELECHANO; STEINMETZ, 2013). Por ser um transcrito *antisense* o *AQP4-AS1* pode ter sua ação na tumorigênese envolvida com alterações da expressão do gene *sense* que codifica a proteína AQP4, porém, mais estudos precisam ser realizados para elucidar as interações desse lncRNA.

No câncer de pulmão foi demonstrada a presença de AQP4 nos tecidos tumorais comparados aos normais não tumorais (XIE *et al.*, 2012). Já no câncer de mama, a redução da expressão da proteína AQP4 foi associada a diminuição da proliferação celular e a alta expressão de seu mRNA foi associada a melhor sobrevida livre recidiva (LI; SUN; HAN, 2016; ZHU, *et al.*, 2019).

Essas informações corroboram com nossas hipóteses sobre a possível interação do rs527616 com a diminuição da expressão da proteína AQP4, que estaria também associada a uma maior suscetibilidade ao câncer de mama.

A identificação de SNPs que interfiram na função ou a expressão gênica e contribuam na suscetibilidade ao câncer de mama são importantes na predição de riscos individuais, populacionais e para a compreensão dos mecanismos de desenvolvimento da doença (LIAN; WANG; ZHANG, 2012), podendo contribuir para melhor eficiência no diagnóstico e na terapia com repercussões prognósticas.

Como perspectiva, seria interessante aumentar o número de pacientes do subtipo triplo-negativo para verificar se os genótipos GG/CG têm uma maior prevalência dentro deste grupo, por se tratar do subtipo mais agressivo.

7. CONCLUSÃO

Pela análise do banco de dados TCGA e estudo GWA, obtivemos 14 lncRNAs diferencialmente expressos nos subtipos de câncer de mama com SNPs associados a essa neoplasia com o $MAF > 0.2$. Destes selecionamos o rs527616 no lncRNA *AQP4-AS1* para análises genotípicas em uma população de mulheres de Curitiba/PR.

Descrevemos pela primeira vez em uma população brasileira, que o polimorfismo rs527616 (C>G) está positivamente associado com a suscetibilidade ao câncer de mama, sendo os genótipos GG e CG, os de risco para a doença.

Como a distribuição alélica não apresentou variação estatisticamente significativa, sugerimos, pelos resultados das análises realizadas, que o risco de susceptibilidade a doença esteja associado ao genótipo e não ao alelo.

Mais estudos com amostras mais consistentes são necessários para avaliar os subtipos de câncer de mama com diferentes comportamentos de agressividade e o rs527616.

8. REFERÊNCIAS

AGRE, P. KING, L.S. YASUI, M, *et al.* Aquaporin water channels--from atomic structure to clinical medicine. **Journal of Physiology**, v. 542.1, p. 3-16. 2002

BATEMAN, A. C.; SHAW, E. C. Breast pathology. **Surgery (United Kingdom)**, v. 34, n. 1, p. 1–7, 2016. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.mpsur.2015.10.002>>.

BEIGUELMAN, B. Curso Prático de Bioestatística. Ribeirão Preto: **Revista Brasileira de Genética**, p.115-117. 1988.

BÉJAR. M.O; MAS, A.M; GONZÁLEZ, J; *et al.* The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. **Genome Biol**, v.18. 2017.

CALLE, A.S; KAWAMURA, Y; YAMAMOTO, Y; *et al.* Emerging roles of long non-coding RNA in cancer. **Cancer Science**. 2018.

CASTLE, N.A. Aquaporins as targets for drug discovery. **Drug Discovery Today**, v. 10, n. 7. 2005.

CIPOLLA, G. A; DE OLIVEIRA, J. C; SALVIANO-SILVA, A; *et al.* Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity. **Noncoding RNA**, v. 4. 2018.

COWPER-SAL·LARI, R.; ZHANG, X.; WRIGHT, J.B.; *et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. **Nature Genetics**. v. 44, p. 1191–1198. 2012.

DAJANI, S. SARIPALLI, A, SHARMA-WALIA. N. Water transport proteins—

aquaporins (AQPs) in cancer biology. **Oncotarget**, v. 9, n. 91, p. 36392-36405. 2018.

DE OLIVEIRA, J. C; OLIVEIRA, L. C; MATHIAS, C; *et al.* Long non-coding RNAs in cancer: Another layer of complexity. **J Gene Med**, v. 21. 2019.

DING, T.; MA, Y.; LI, W.; *et al.* Role of aquaporin-4 in the regulation of migration and invasion of human glioma cells. **Int J Oncol**, v. 38, p. 1521-1531. 2011

DING, T.; ZHOU, Y.; SUN, K.; *et al.* Knockdown a Water Channel Protein, Aquaporin-4, Induced Glioblastoma Cell Apoptosis. **PLoS One**. 2013.

EROLE, P.; BOSCH, A.; ALEJANDRO PÉREZ-FIDALGO, J.; LLUCH, A. Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. **Cancer Treatment Reviews**, v. 38, n. 6, p. 698–707, 2012. Elsevier Ltd. Disponível em: <http://dx.doi.org/10.1016/j.ctrv.2011.11.005>.

FENG, H; ZHANG, J; SHI, Y; *et al.* Long noncoding RNA LINC-PINT is inhibited in gastric cancer and predicts poor survival. **Journal of Cellular Biochemistry**. v. 120, p. 9594-9600. 2019.

FERLAY, J.; SOERJOMATARAM I, I.; DIKSHIT, R.; *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. **Int J Cancer**, v. 136, n. 5, p. E359-386, 2015. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/25220842>.

GEISLER, S.; COLLIER, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. **Nature Reviews Molecular Cell Biology**, v. 14, n. 11, p. 699–712, 2013. Disponível em: <http://www.nature.com/doi/10.1038/nrm3679>.

GOLDHIRSCH, A. *et al.* Personalizing the treatment of women with early breast cancer: Highlights of the st gallen international expert consensus on the primary therapy of early breast Cancer 2013. **Annals of Oncology**, v. 24, n. 9, p. 2206–2223, 2013.

GUDMUNDSDOTTIR, E. T.; BARKARDOTTIR, R. B.; ARASON, A.; *et al.* The risk allele of SNP rs3803662 and the mRNA level of its closest genes TOX3 and LOC643714 predict adverse outcome for breast cancer patients. **BMC Cancer**, v. 12, p. 1-10. 2012.

GWAS CATALOG. **The NHGRI-EBI Catalog of published genome-wide association studies.** 2018. Disponível em: <https://www.ebi.ac.uk/gwas/anniversary>.

HALLADAY, J. R.; LENHART, K. C.; ROBASKY, K., *et al.* Applicability of Precision Medicine Approaches to Managing Hypertension in Rural Populations. **Journal of Personalized Medicine**, v. 8, p. 1-28. 2018.

HE, JH.; HAN, ZP.; LIU, JM.; *et al.* Overexpression of long non-coding RNA *MEG3* inhibits proliferation of hepatocellular carcinoma Huh7 cells via negative modulation of miRNA-664. **Journal of Cellular Biochemistry**. 2017.

HENG, Y. J.; LESTER, S. C.; TSE, G. M. K.; *et al.* The molecular basis of breast cancer pathological phenotypes. **Journal of Pathology**, v. 241, n. 3, p. 375–391, 2017.

HORTOBAGYI, G. N.; SALAZAR, J. G.; PRITCHARD, K.; *et al.* The Global Breast Cancer Burden: Variations in Epidemiology and Survival. **Clinical Breast Cancer**, v. 6, n. 5, p. 391-401. 2005.

INSTITUTO NACIONAL DE CANCER JOSÉ ALENCAR GOMES DA SILVA.

Estimativa 2020: incidência de câncer no Brasil / Instituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro: **INCA**, 2019.

KIM, S; MISRA, A. SNP Genotyping Technologies and Biomedical Applications. **Annu. Rev. Biomed. Eng.** p. 289-320. 2007.

KREIDA, S; HORSEFIELD, S.T. Structural insights into aquaporin selectivity and regulation. **Current Opinion in Structural Biology**, v.33, p.126–134. 2015.

KWOK, Z. H.; TAY, Y. Long noncoding RNAs: links between human health and disease. **Biochemical Society Transactions**, v. 45, n. 3, p. 805–812, 2017. Disponível em: <<http://biochemsoctrans.org/lookup/doi/10.1042/BST20160376>>.

LEE, S.J; CHAE, Y.S; KIM, W.W; et al. AQP5 expression predicts survival in patients with early breast cancer. **Annals of Surgical Oncology**. v. 21, p. 375-383. 2014.

LI, Y. B.; SUN, S. R.; HAN, X. H. Down-regulation of AQP4 Inhibits Proliferation, Migration and Invasion of Human Breast Cancer Cells. **Folia Biologica**, v. 62, p. 131-137. 2016.

LI, J., HAN, L.; ROEBUCK, P.; DIAO, L.; LIU, L.; YUAN, Y.; WEINSTEIN, J. N.; LIANG, H.; TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. **Cancer Research**, v. 75, p. 3728-3737. 2015.

LI, J.; LIU, R.; TANG. S.; *et al.* The effect of long noncoding RNAs HOX transcript antisense intergenic RNA single-nucleotide polymorphisms on breast cancer, cervical cancer, and ovarian cancer susceptibility: A meta-analysis. **J Cell Biochem**, p. 1-12. 2018.

LI, L. ZHANG, H; et al. Proinflammatory role of aquaporin-4 in autoimmune neuroinflammation. **The FASEB Journal**. v. 25, p. 1556-1566. 2011.

LI, L; ZHANG, G.Q; CHEN, H; *et al.* Plasma and tumor levels of Linc-pint are diagnostic and prognostic biomarkers for pancreatic cancer. **Oncotarget**. v. 7, p. 71773-71781. 2016.

LI, R; ZHANG, L; QUI, Z; *et al.* High LINC00536 expression promotes tumor progression and poor prognosis in bladder cancer. **Experimental Cell Research**, p. 32-40. 2019.

LIAN, H; WANG, L; ZHANG, J. Increased Risk of Breast Cancer Associated with CC Genotype of Has-miR-146a Rs2910164 Polymorphism in Europeans. **PLoS ONE**, v. 7, n. 2, p. e31615. 2012.

LIU, Y.; SHARMA, S.; WATABE, K. Roles of lncRNA in breast cancer. **Front Biosci**, v. 7, p. 94-108. 2017.

LIU, R; WANG, L; GAN, T; *et al.* Long noncoding RNA LINC00511 promotes cell growth and invasion in triple-negative breast cancer by interacting with Snail. **Cancer Management and Research**, v. 11, p. 5691-5599. 2019.

LOSKO, M; KOTLINOWSKI, J; JURA, J. Long Noncoding RNAs in Metabolic Syndrome Related Disorders. **Mediators of Inflammation**, 2016.

LU, K.H; LI, W; LIU, X.H; *et al.* Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. **BMC Cancer**. 2013.

MAKKI, J. Diversity of breast carcinoma: Histological subtypes and clinical relevance. **Clinical Medicine Insights: Pathology**, v. 8, n. 1, p. 23–31, 2015.

MATTICK, J. S. Non-coding RNAs: The architects of eukaryotic complexity. **EMBO Reports**, v. 2, n. 11, p. 986–991, 2001.

MATTICK, J. S. The genetic signatures of noncoding RNAs. **PLoS Genetics**, v. 5, n. 4, 2009.

MICHAILIDOU, K.; LINDSTROM, S.; BEESLEY, J; *et al.* Association analysis identifies 65 new breast cancer risk loci. **Nature**, 551(7678), p. 92–94, 2017.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION – NCBI. U.S. **National Library of Medicine**. 2019. Disponível em: <
<https://www.ncbi.nlm.nih.gov/>>

NING, S; ZHANG, Y. Genome-Wide Mapping of SNPs in Non-coding RNAs. **Advances in Experimental Medicine and Biology**. v. 1094, p. 39-48. 2018.

OUYANG, W; REN, L; LIU, G; *et al.* LncRNA MIR4435-2HG predicts poor prognosis in patients with colorectal cancer. **PeerJ**. 2019.

PELECHANO, V; STEINMETZ, L. M. Gene Regulation by Antisense Transcription. **Nature Reviews**, v. 14, p. 880-893. 2013.

PEROU C. M.; SORLIE. T; EISEN. M. B; *et al.* Molecular portraits of human breast tumours. **Nature**, v. 406. n. 6797. p. 747-752, 2000.

PRAT, A; PARKER, JS; KARGINOVA, O; *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. **Breast Cancer Res**. 2010.

PROBST, C. M. *et al.* HLA polymorphism and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Paraná, Brazil. **Human biology**, v. 72, n. 4, p. 597–617, 2000.

QIAN, H; CHEN, L; HUANG, L; The lncRNA MIR4435-2HG promotes lung cancer progression by activating β -catenin signalling. **Journal of Molecular Medicine**, v. 96, p. 753-764. 2018.

QIN, R.; CHEN, Z.; DING, Y.; *et al.* Long non-coding RNA *MEG3* inhibits the proliferation of cervical carcinoma cells through the induction of cell cycle arrest and apoptosis. **Neoplasma**, v. 60, p. 486-492. 2013.

SANGER, F; NICKLEN, S; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology*. v.24, p. 104-108. 1992.

SHI, Z; ZHANG, T. LOU, L. *et al.* Aquaporins in human breast cancer: identification and involvement in carcinogenesis of breast cancer. **J Surg Oncol**, v. 206, p. 267-272. 2011.

SHENG, X.; LI, J.; YANG, L.; *et al.* Promoter hypermethylation influences the suppressive role of maternally expressed 3, a long non-coding RNA, in the development of epithelial ovarian cancer. **Oncology Reports**, v. 32, p.277-285. 2014.

SORLIE, T.; TIBSHIRANI, R.; PARKER, J.; *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. **Proceedings of the National Academy of Sciences**, v. 100, n. 14, p. 8418–8423, 2003. Disponível em: <http://www.pnas.org/cgi/doi/10.1073/pnas.0932692100>.

SUN, M.; XIA, R.; JIN, F. *et al.* Downregulated long noncoding RNA *MEG3* is associated with poor prognosis and promotes cell proliferation in gastric cancer. **Tumor Biology**, v. 35, p.1065-1073. 2014.

TIAN, ZZ.; GUO, XJ.; ZHAO, YM.; FANG, Y. Decreased expression of long non-coding RNA *MEG3* acts as a potential predictor biomarker in progression and poor

prognosis of osteosarcoma. **International Journal of Clinical and Experimental Pathology**, v. 8, p. 15138-15142. 2015.

THE CANCER GENOME ATLAS – TCGA. **The website of the National Cancer Institute**. 2018. Disponível em: <<https://cancergenome.nih.gov>>.

TRADTRANTIP, L; JIN, B; YAO, X; et al. Aquaporin-Targeted Therapeutics: State-of-the-Field. **Advances in Experimental Medicine and Biology**. v. 969, p. 239-250. 2017.

UPADHYAYA, A; SMITH, R.A; CHACON-CORTES, D; *et al.* Association of the microRNA-Single Nucleotide Polymorphism rs2910164 in miR146a with sporadic breast cancer susceptibility: A case control study. **Gene**, v. 576, p. 256-260. 2016.

VANDERBROEK, A; YASUI, M. Regulation of AQP4 in the Central Nervous System. **Internation Journal of Molecular Sciences**. v. 21. 2020.

VERKMAN, A. S.; HARA-CHIKUMA, M.; PAPADOPOULOS, M. C. Aquaporins—new players in cancer biology. **Journal of Molecular Medicine**, v. 86, p. 523-529. 2008.

VERKMAN, A.S; BINDER. D.K; *et al.* Three distinct roles of aquaporin-4 in brain function revealed by knockout mice. **Biochim Biophys Acta**. 2006

VISSCHER, P. M; WRAY, N. R; ZHANG, Q; *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. **The American Journal of Human Genetics** v. 101, p. 5–22. 2017.

WANG, R.P; JIANG, J; JIANG T; *et al.* Increased long noncoding RNA LINC00511 is correlated with poor prognosis and contributes to cell proliferation and metastasis by modulating miR-424 in hepatocellular carcinoma. **European Review for Medical**

and Pharmacological Sciences, v. 23, p. 3291-3301. 2019.

WANG, J; TIAN, Y; ZHENG, G; *et al.* An integrated analysis reveals the oncogenic function of lncRNA LINC00511 in human ovarian cancer. **Cancer Medicine**. v. 8, p. 3026-3035. 2019.

WARTH, A; MULEY, T; MESTER, M; *et al.* Loss of aquaporin-4 expression and putative function in non-small cell lung cancer. **BioMed Central**. 2011.

WILEY-BLACKWELL. TNM Classification of Malignant Tumours. **UICC International Union Against Cancer**. Seventh Edition, 2009.

WORLD HEALTH ORGANIZATION. **International Agency for Research on Cancer**. Globocan 2018.

WU, T.; DU, Y. LncRNAs: From basic research to medical application. **International Journal of Biological Sciences**, v. 13, n. 3, p. 295–307, 2017.

WYLIE, A. A.; MURPHY, S. K.; ORTON, T. C.; JIRTLE, R. L. Novel Imprinted DLK1/GTL2 Domain on Human Chromosome 14 Contains Motifs that Mimic Those Implicated in IGF2/H19 Regulation. **Genome Research**, p. 1711-1718. 2000.

XIE, Y.; WEN, X.; JIANG, Z.; *et al.* Aquaporin 1 and aquaporin 4 are involved in invasion of lung cancer cells. **Clin Lab**, v. 58, p. 75-80. 2012.

XING, C.; CAI, Z.; GONG J.; *et al.* Identification of Potential Biomarkers Involved in Gastric Cancer Through Integrated Analysis of Non-Coding RNA Associated Competing Endogenous RNAs Network. **Clin Lab**. v. 64, p. 1661-1669. 2018.

YAN, L; WU, X; LIU, Y; XIAN, W. LncRNA Linc00511 promotes osteosarcoma cell proliferation and migration through sponging miR-765. **Journal of Cellular**

Biochemistry. 2018.

YANG, G; LU, X; YUAN, L. LncRNA: A link between RNA and cancer. **BBA - Gene Regulatory Mechanisms**, 2014.

ZHANG, B; LI, C; SUN, Z. Long non-coding RNA LINC00346, LINC00578, LINC00673, LINC00671, LINC00261, and SNHG9 are novel prognostic markers for pancreatic cancer. **American Journal of Translational Research**, v. 10, p. 2648-2658. 2018.

ZHOU, Y.; ZHONG, Y.; WANG, Y.; *et al.* Activation of p53 by *MEG3* non-coding RNA. **The Journal of Biological Chemistry**. v. 282, n. 34, p. 24731–24742. 2007.

ZHU, L; MA, N; WANG, L; *et al.* Significant prognostic values of aquaporin mRNA expression in breast cancer. **Cancer Management and Research**. v.11, p. 153-1515. 2019.

ZHU, Y.; CHEN, P.; GAO, Y.; *et al.* *MEG3* Activated by Vitamin D Inhibits Colorectal Cancer Cells Proliferation and Migration via Regulating Clusterin. **EBioMedicine**, v. 30, p. 148-157. 2018.